Expression of rat liver ketohexokinase in yeast results in fructose intolerance

Iain A. DONALDSON,* Timothy C. DOYLE and Nada MATAS†
Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

Rat liver ketohexokinase (ATP:D-fructose 1-phosphotransferase; EC 2.7.1.3) was purified to homogeneity and the molecular mass of the protein was found by mass spectrometry to be 32800 Da. The enzyme was cleaved and the amino acid sequences of seven peptides, comprising 24% of the total sequence, were determined. This sequence information was used to design oligonucleotide primers for a PCR using rat liver single-stranded cDNA as a template. The 224 bp PCR product was used as a probe to screen a rat liver cDNA library. A cDNA sequence of 1342 bp was obtained from three positive clones. This contained the entire coding region for ketohexokinase, and all seven peptides were identified in the predicted amino acid sequence. When ketohexokinase was expressed in *Saccharomyces cerevisiae* using the yeast expression vector pMA91, the cells became intolerant of the presence of fructose in their growth media. The growth of an exponential-phase culture was completely arrested within 90 min by the addition of fructose to a final concentration as low as 0.1% (w/v). This response is associated with an accumulation of fructose 1-phosphate. The cDNA for ketohexokinase encodes a protein composed of 299 amino acids with a combined molecular mass of 32728 Da. This is in close agreement with the value for the isolated protein determined by mass spectrometry. The primary structure does not show any significant homology with those of other eukaryotic hexokinases, but it contains a highly conserved region that is present in three prokaryotic phosphotransferases that have furanose substrates.

INTRODUCTION

Ketohexokinase (ATP:D-fructose 1-phosphotransferase; EC 2.7.1.3) is the enzyme responsible for the primary metabolism of dietary fructose in mammals [1,2]. The enzyme is most abundant in the liver, but is also present in the kidney, small intestine and pancreas [3]. All of these tissues possess a specialized pathway employing ketohexokinase, aldolase B (EC 4.1.2.13) and triokinase (EC 2.7.1.28), which together convert fructose into intermediates for glycolysis and gluconeogenesis. Two inherited abnormalities of this pathway have been recognized in humans. Hereditary fructose intolerance is a potentially lethal condition caused by a deficiency of aldolase B [4]. As a result of the activity of ketohexokinase, ingestion of fructose by individuals with this disorder is followed by accumulation of fructose 1-phosphate and depletion of ATP in tissues possessing this pathway (liver, kidneys, small intestine and pancreas). Essential fructosuria, in contrast, is a benign defect caused by the absence of ketohexokinase, which results in an inability to synthesize fructose 1-phosphate [5]. Under these circumstances, uptake of dietary fructose leads to a transient increase in the concentration of blood fructose. This gradually decreases, and approx. 80–90% of the ingested fructose is eventually metabolized, the rest being excreted in urine, indicating that, under these circumstances, fructose is being utilized by other pathways [2].

Mammalian hexokinases (EC 2.7.1.1) are also able to phosphorylate fructose [6]. In rats there are four isoforms of this enzyme, belonging to a family of kinases that display antigenic cross-reactivity and regions of amino acid homology [7]. Brain contains hexokinase type I only, skeletal muscle contains type II, and adipose tissue and heart contain both of these. Kidney contains types I and III, whereas the intestine and liver contain isoenzymes I–III. Isoenzymes I, II and III have affinities for glucose roughly ten times those for fructose, phosphorylate a range of other hexoses *in vitro*, and are strongly inhibited by glucose 6-phosphate [8]. Hexokinase type IV (glucokinase; EC 2.7.1.1), the predominant isoenzyme in the liver, is not inhibited by its product, and is regulated by insulin and glucagon [6,8]. It shows a similar substrate specificity to the other hexokinases, but has a much lower affinity for its substrates, and is not responsible for the primary metabolism of fructose *in vivo*. Nonetheless, this enzyme has recently been shown to have substantial regions of amino acid homology with isoenzymes I–III, and so was confirmed to be a member of the hexokinase family, which also includes the yeast isoenzymes PI and PII [7].

The properties of ketohexokinase show substantial differences from those of enzymes belonging to the hexokinase family. It phosphorylates its substrate at the C-1 position rather than the C-6 position [9,10], has an absolute requirement for monovalent cations (physiologically, K^+) [11], will only phosphorylate furanose substrates and, unlike glucokinase, is not induced hormonally or by its substrate. In common with hexokinases, it does show marked preference for Mg^{2+}-ATP as a nucleotide, and is inhibited by ADP.

Liver ketohexokinase has been isolated from a number of sources [10,12–14]. Extensive studies of the enzyme's kinetic properties and substrate specificity have been performed. However, a primary structure of ketohexokinase has not been reported.

Saccharomyces cerevisiae has been used to express a number of

Abbreviations used: TEA, triethanolamine; PTH-amino acid, phenylthiohydantoin-amino acid; TFA, trifluoroacetic acid.

* To whom correspondence should be addressed.
† Permanent address: Department of Molecular Biology, Faculty of Natural Sciences and Mathematics, University of Zagreb, Rooseveltov trg 6, P.O. Box 933, YU-41001 Zagreb, Croatia.

The nucleotide sequence data reported in this paper have been submitted to the Genbank and EMBL Data Banks under accession numbers M86235 and X63658 respectively.

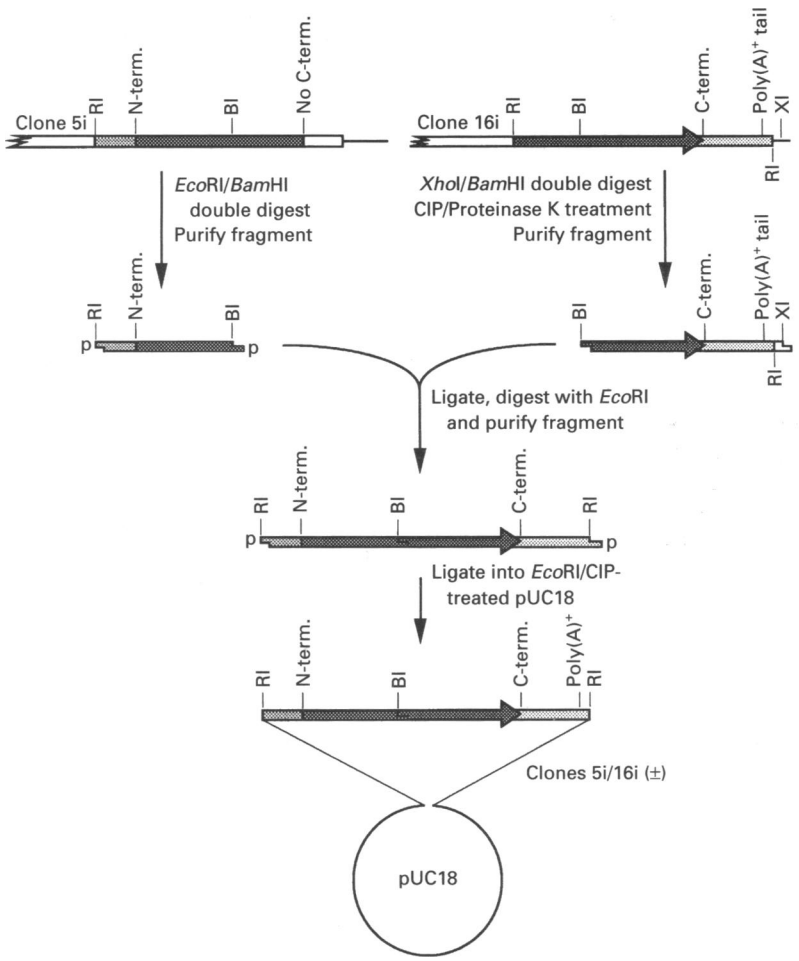

Figure 1 Strategy for the construction of hybrid clone 5i/16i in pUC18

The co-ligated cDNA was removed from the ketohexokinase cDNA at the *Eco*RI sites (RI), and the 5i and 16i clones were spliced at the *Bam*HI site (BI). The clone 16i was originally cut with *Xho*I (XI) and *Bam*HI, before treatment with calf intestinal phosphatase (CIP). The heavily shaded boxes represent the ketohexokinase coding region, the lightly shaded boxes represent 5′ or 3′ untranslated regions, the open boxes represent unassociated, co-ligated cDNA, and the thin lines represent plasmid DNA. C-term and N-term indicate C- and N-termini respectively.

active mammalian proteins [15]. Since yeast does not contain aldolase B, it was of interest to us to investigate whether or not the expression of active ketohexokinase in this organism would condition fructose intolerance. Such a strain would be of great interest to us for selection of fructotransferase mutants of invertase.

We report the cloning and complete nucleotide sequence for the cDNA of rat liver ketohexokinase and the effects that the expression of this enzyme in yeast has upon the response of the yeast to the provision of fructose in the growth medium.

EXPERIMENTAL

Purification of rat liver ketohexokinase

Eight male Sprague–Dawley rats (250–300 g) were killed by cervical dislocation and exsanguinated by severing the carotid arteries. Their livers were used immediately for enzyme purification.

The procedure used to purify ketohexokinase was a modification of previously published methods [10,12–14], using improved column materials and conditions. Gel filtration was performed on a 108 cm × 4 cm^2 Sephacryl S100 (Pharmacia LKB Biotechnology) column eluted with triethanolamine (TEA) buffer (10 mM TEA/HCl, pH 7.3, 1 mM dithiothreitol) at a flow rate of 3 ml/min. Anion-exchange chromatography was performed on a 19 cm × 5 cm^2 Fractogel TSK DEAE-650(S) (Merck) column eluted with a 0–200 mM KCl gradient in TEA buffer at 2 ml/min. Chomatofocusing was performed on a 15 cm × 0.75 cm^2 PBE94 column (Pharmacia LKB Biotechnology) eluted at a flow rate of 0.33 ml/min with Polybuffer 74/HCl, pH 4.5 (Pharmacia LKB Biotechnology).

Enzyme activity in fractions of column eluent was assayed as described below, and the activity of pooled samples was measured by the method of Adelman et al. [12]. For the colorimetric assay, a 5 μl aliquot of sample was added to 1 ml of reaction mixture (75 mM potassium phosphate, pH 7.2, 300 mM KCl, 20 mM $MgCl_2$, 15 mM ATP, 5 mM phospho*enol*pyruvate, 0.2 mM TPP, 8 μM FAD, 10 mM phenol, 8 mM 4-aminoantipyrene and 5 mM fructose) containing 2 units of peroxidase, 2 units of pyruvate oxidase (Boehringer Mannheim) and 3 units of pyruvate kinase (Sigma). The reaction was incubated at 37 °C for 60 min prior to quenching with 100 μl of 110 mM EDTA. The absorbance at 510 nm was then read spectrophotometrically. Protein concentrations were determined by the method of Bensadoun and Weinstein [16]. SDS/PAGE was performed according to Hames [17].

The molecular mass of the purified protein was accurately determined by matrix-assisted laser adsorption on a Finnigan MAT LASERMAT 'time-of-flight' mass spectrometer [18]. A 0.2 μl or 1.0 μl sample of the purified protein (approx. 800 μg/ml) was mixed with 0.2 μl of sinapinic acid matrix on a gold-plated stainless steel target and allowed to dry under an infra-red lamp. This was then put in the path of the ionizing laser of the mass spectrometer and the relative molecular masses of ions generated during irradiation were measured.

Preparation and sequencing of peptides of ketohexokinase

A sample (approx. 30 μg) of purified ketohexokinase in 45 μl of 100 mM ammonium carbonate/acetic acid buffer, pH 8.1, was diluted with 50 μl of 100 mM ammonium bicarbonate, pH 8.3, containing 5% (v/v) acetonitrile. To this, 2 μg each of endoproteinase Glu-C and trypsin (Boehringer, Sequencing Grade) were added and the reaction was incubated for 60 min at 37 °C. A control reaction containing proteinase only was also performed. After digestion, 100 μl of 8 M guanidine hydrochloride/0.5 M Tris/0.35 mM EDTA was added to each reaction. The mixtures were then reduced by the addition of 2 μl of a 100 mg/ml solution of dithiothreitol and incubated at 37 °C for 60 min. In order to alkylate the cysteine residues, a 2 μl volume of 4-vinylpyridine was then added to each reaction and the mixtures were incubated at 37 °C for 30 min. Both samples were then chromatographed by reverse-phase h.p.l.c. on a Brownlee Aquapore RP300 (C8) 220 mm × 2 mm column, loading with 0.1% (v/v) trifluoroacetic acid (TFA) in water and then applying a 5–50% (v/v) gradient of 0.09% (v/v) TFA in 80% (v/v) acetonitrile over 45 min, followed by a 50–90% (v/v) gradient over 15 min. The sample and control were then compared to eliminate peptides derived from the proteinases themselves, and the six major peaks were sequenced on an Applied Biosystems 470A protein sequencer with on-line phenylthiohydantoin (PTH) analysis.

Cloning of the ketohexokinase probe

Unless specifically mentioned, all procedures were performed according to Sambrook et al. [19].

Preparation of PCR primers

Because the relative positions of peptides KHK 45 and 49 (see Figure 3) were not known, four oligonucleotide primers were synthesized to account for both possible orientations (primer 1, ATACTRTGDTTRGASGGWCTACA; primer 2, TATGAYACHAAYCTSCCWGATGT; primer 3, AARGTYGASAAACCBATACCWCT; primer 4, TTYCARCTSTTTGGVTATGGWGA, using the letter codes recommended by the IUPAC/IUB Biochemical Nomenclature Commission). The primers were prepared on an Applied Biosystems DNA synthesizer and purified on oligonucleotide purification cartridges (Applied Biosystems 400771). The concentrations of the oligonucleotides were determined spectrophotometrically.

Amplification and cloning of the ketohexokinase probe from a single-stranded cDNA template

A rat liver was dissected as described above. Isolation of RNA from 5 g of rat liver tissue was performed by directly scaling up the acid guanidinium thiocyanate/phenol/chloroform (AGPC) extraction method of Chomczynski and Sacchi [20]. Single-stranded cDNA was prepared from 20 μg of crude RNA by employing the method for purified mRNA described in section 8.60–8.61 of Sambrook et al. [19] using Moloney murine leukaemia virus reverse transcriptase (Gibco BRL) and oligo$(dT)_{18-20}$ (Pharmacia LKB Biotechnology) as a primer. PCR was performed directly on 20 μl of this cDNA reaction mixture as described in section 7.26–7.29 of Sambrook et al. [19], using oligonucleotide primers 1 and 4 or 2 and 3. Single primer control reactions were also performed for all oligonucleotides, and three $MgCl_2$ concentrations (1.5 mM, 3.0 mM and 5.0 mM) were used for all reactions. Thirty cycles of PCR were carried out on a Perkin–Elmer Cetus DNA thermal cycler. Cycles were: 94 °C, 1.5 min; 50 °C, 2 min; 72 °C, 2 min (10 min for the last cycle). Cloning of the PCR products into pUC18 [21] was performed by the method of Bonthron [22], followed by transformation into *Escherichia coli* strain DH5α. The nucleotide sequence of both strands was determined by the dideoxy method [23] using the Pharmacia T7 Sequencing Kit (27-1682-01) employing deoxyadenosine 5′-[α-^{35}S]thiotriphosphate (Amersham).

Labelling of the probe

The ketohexokinase fragment was excised from pUC18 and used for the synthesis of a mixture of randomly primed probes, labelled with [^{32}P]dCTP (3000 Ci/mmol) (Amersham), by the method described in section 10.14–10.17 of Sambrook et al. [19]. The mixture of labelled probes was purified from the reaction mixture using Nensorb 20 columns (NEN Research Products).

Screening of a cDNA library in λZAP

The rat liver cDNA library was kindly provided by Professor Brian Burchell, University of Dundee. It was prepared from Sprague–Dawley rat liver mRNA and cloned into λZAP [24] with a phage titre of 10^9 plaque-forming units/ml. The R408 f1 helper phage and *E. coli* strains XL1-Blue and BB4 were gifts from Dr. Tamar Enoch (Microbiology Unit, University of Oxford). Confluent plates with BB4 cells as hosts were blotted in duplicate with Hybond N+ nylon filters (Amersham), and treated according to sections 2.112–2.113 of [19], substituting 100 mM sodium phosphate, pH 7.2, for the 0.5 M Tris/HCl used in the neutralizing solution. Hybridization was performed according to sections 2.114–2.117 of [19], substituting 100 mM sodium phosphate (pH 7.2)/1% (w/v) SDS/1 M NaCl for the SSC, and blocking with 200 μg/ml salmon sperm DNA and 5 × Denhart's solution (section 9.49 [19]). The filters were then washed five times at high stringency with 10 mM sodium phosphate, pH 7.2, and 1% (w/v) SDS at 65 °C and allowed to dry before autoradiography. Clones hybridizing the probe were isolated by successive replating and screening. Single clones were finally suspended in 1 ml of 100 mM NaCl, 8 mM $MgSO_4$, 0.01% (w/v) gelatin and 50 mM Tris/HCl, pH 7.5 (section A.7 [19]).

The sizes of the inserts in the clones were determined by amplifying this region by PCR using the M13 forward and reverse primers flanking the cloning site of λZAP [25]. Thirty cycles of PCR were performed as described above. The DNA was separated on a 1% agarose gel, Southern-blotted (section 9.31–9.46 in [19]) and the membrane probed at high stringency with the ^{32}P-labelled DNA probe. An autoradiograph revealed the size of the clonal inserts. Clones were rescued into the phagemid pBluescript SK using R408 helper phage, which was used to transform *E. coli* XL1-Blue [24].

Rescued clones were mapped using the restriction enzymes *Apa*I, *Bam*HI, *Eco*RI, *Hin*dIII, *Kpn*I, *Not*I, *Pst*I, *Sma*I, *Sst*I, *Xba*I and *Xho*I (Gibco BRL), unique cleavage sites for which exist within the multiple cloning site of the pBluescript SK. Approx. 300 bp of nucleotide sequence of each of the clones was determined using the M13 forward and reverse primers.

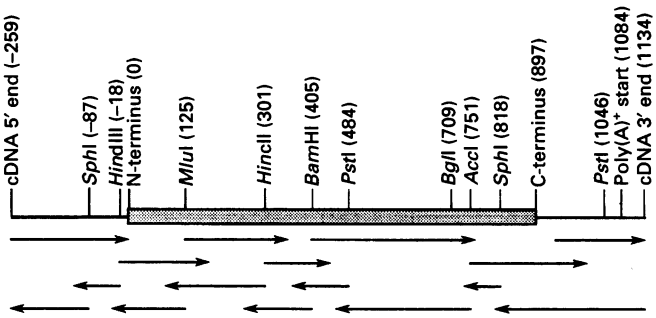

Figure 2 Sequencing strategy for the complete cDNA of ketohexokinase

The arrows represent the regions of the gene sequenced by each deleted clone. The restriction enzyme sites of importance within the cDNA are marked, along with their position (in bp) from the first nucleotide in the start codon.

Unidirectional deletion of clone 16i by exonuclease treatment was performed by a modification of the method described in section 13.34–13.41 of [19], substituting mung-bean nuclease for S1 nuclease. The clone was cut with *Xho*I and *Kpn*I and digested with *Exo*III according to the published method. To each of the aliquots taken, 10 × concentrated mung-bean buffer [22] and mung-bean nuclease to 1 unit/μl were added. Mixtures were then incubated at room temperature for 20 min and quenched with 6 × stop solution (800 mM Tris/HCl, pH 8.0, 20 mM EDTA, 80 mM $MgCl_2$). The samples were then analysed and treated as in the published method [19] and the nucleotide sequences were determined as described above.

Clones 16i and 5i were cut and re-ligated according to Figure 1 to produce a clone (16i/5i) containing the coding region and the 3′ and 5′ untranslated regions. Clones were isolated which contained the hybrid ketohexokinase gene inserted into the plasmid in both directions. The co-ligated, unassociated DNA was removed from clone 5ii by cleaving with *Eco*RI and ligating the fragment containing ketohexokinase cDNA into pUC18 to give clone 5ii/a. Endonuclease digestion and re-ligation of each orientation of the two clonal inserts 16i/5i and 5ii/a yielded a set of deleted clones, enabling the nucleotide sequence on both strands to be determined using the M13 forward and reverse primers (Figure 2).

Expression of ketohexokinase in *Saccharomyces cerevisiae*

The coding region of the ketohexokinase cDNA was amplified from clone 5ii/a by PCR using oligonucleotide primers with the following sequences: upstream primer, 5′ TTGGAGATCTA-CACAATGGAAGAGAAGCAGAT 3′; downstream primer 5′ CCCACCGCAGATCTCTCACACAATGCCATCAA 3′. Both primers incorporate non-complementary *Bgl*II sites and the upstream primer also contains the yeast consensus sequence [26] immediately 5′ to the start codon. The start and stop codons in the sequences are underlined. The conditions used for PCR were as described above, using 1 mM $MgCl_2$, and the 911 bp product was blunt-end-ligated into the Klenow-treated *Sph*I site of pUC18. A 693 bp *Mlu*I/*Sph*I fragment was excised from the centre of the cloned PCR product and replaced with the equivalent fragment from clone 5ii/a. The resulting clone was named pUC KHK-G_7. In order to check for any residual PCR errors, the nucleotide sequences of the remaining PCR-generated flanking segments were confirmed using primers 1 and 4 in the sequencing reactions.

The coding region for ketohexokinase was excised from pUC KHK-G_7 with *Bgl*II and ligated into the *Bgl*II cloning site of the yeast expression vector pMA91 [27]. Clones in the correct orientation for expression were termed pMA91-KHK and those in the opposite orientation were termed pMA91-KHK(−). These plasmids were used to transform *Saccharomyces cerevisiae* strain TDSII (*leu2-3, leu2-112, trp1, suc°, mal°, hxk1, hxk2, GLK, pho8-2, pho3Δ1, pho5Δ1*, MATα) using a modification of the method of Ito et al. [28]. The TDS II cells were prepared essentially as published, using lithium acetate, but the transformation mixture contained 100 μl of cell suspension, 10 μl of carrier DNA (sonicated salmon sperm DNA; 25 mg/ml), 1 μg of plasmid DNA and 10 μl of absolute ethanol. This mixture was incubated for 30 min at 30 °C before the addition of 1 ml of 40 % (w/v) poly(ethylene glycol) 4000, 100 mM CH_3CO_2Li, 10 mM Tris/HCl, pH 8.0, and 1 mM EDTA, and the incubation was continued at 30 °C for 1 h. The cells were then pelleted and washed with water before being resuspended in 1 ml of YEPD [1 % (w/v) yeast extract, 1 % (w/v) bactopeptone, 2 % (w/v) glucose] and incubating at 30 °C for 6 h without shaking. This allows the cells to obtain a high copy number of the plasmid. The cells were then plated on selective media as published [28].

The presence of the ketohexokinase gene in the transformants was confirmed by the preparation of yeast plasmid DNA minipreps and by the amplification of the coding region by PCR using the same primers and conditions originally used to generate this fragment.

Intolerance of fructose was demonstrated by replica plating of transformants on to SD Leu^- plates [0.17 % (w/v) yeast nitrogen base without amino acids, 0.5 % (w/v) ammonium sulphate, 1 % (w/v) glucose, 20 μg/ml each of adenine sulphate, uracil, L-tryptophan, L-histidine-HCl, and 2 % (w/v) agar], either with or without 0.5 % (w/v) fructose. The plates were then incubated for at least 48 h at 30 °C. Colonies possessing the plasmid pMA91-KHK showing intolerance were termed TDB I, and those transformed with pMA91-KHK(−) and showing leucine-independence were termed TDB I(−). The effect of various fructose concentrations on exponential-phase cells growing in SD Leu^- liquid culture was also investigated. Early exponential-phase cultures of TDB I and TDB I(−) were divided into 30 ml aliquots in 50 ml tubes, and growth (at 30 °C with shaking) was followed by taking 1 ml samples and measuring A_{600}. After 2 h, fructose was added to various concentrations (0.025–0.5 %, w/v) and the A_{600} measurements were continued for several hours. Metabolites from TDB I cells, treated or not with fructose, were extracted with perchlorate by the method of den Hollander et al. [29]. The presence of fructose 1-phosphate in these extracts was detected by the method of Eggelston [30].

Cell extracts from yeast strains TDB I, TDB I(−) and TDS II were prepared by vortexing a cell suspension with glass beads [31]. The supernatants were partially purified by ammonium sulphate fractionation and heat treatment as described above, and all fractions were tested for the presence of ketohexokinase by the colorimetric methods described above and assayed by the method of Adelman et al. [12].

RESULTS

Purification and characterization of ketohexokinase

Rat liver ketohexokinase was purified to homogeneity, with a specific activity of 11.6 μmol/min per mg of protein. A 166-fold purification, though low when compared with previous reports, resulted from the use of a purer starting homogenate than those methods employing bladed blenders which break open organelles. The mild homogenization technique that we employed allows the sedimentation of intact organelles, producing a cytoplasmic fraction that is minimally contaminated with proteins released

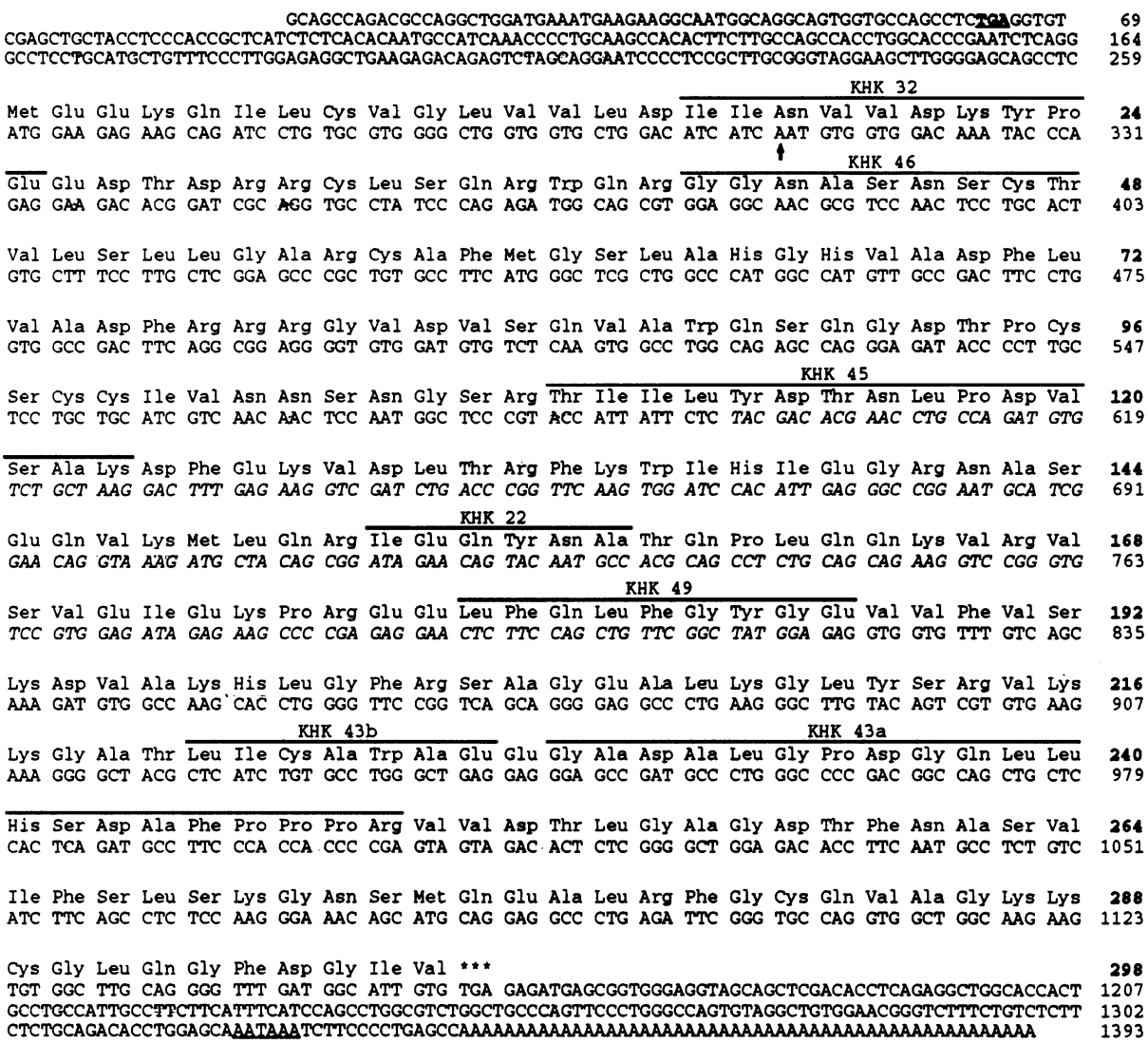

Figure 3 DNA and amino acid sequence of rat liver ketohexokinase

Bases are numbered from the 5′ end of the cDNA sequence in clone 5i. The in-phase termination codon is designated by ***, the in-frame upstream stop codon is in bold underlined letters, and the putative polyadenylation signal is underlined. The region of the structural cDNA that was used to probe the cDNA library is in italics, and the site of the point mutation (A to G) in clone 5i is arrowed. Amino acids are numbered starting with the initiator Met. Amino acids identified by peptide sequencing are indicated by overlines and the peptide name is shown above each line.

from organelles. This is evinced by the considerably higher specific activity of our starting homogenate (0.07 μmol/min per mg of protein) when compared with previous reports [10,12–14]. A 166-fold purification suggests that ketohexokinase constitutes approx. 0.6% of the cytoplasmic protein of rat livers.

The purified protein displayed a relative mobility on SDS/PAGE consistent with a molecular mass of 31 kDa, but a more accurate value of 32.8 kDa was obtained by mass spectrometry. The mass spectrum also revealed a K^+ peak, which, since the buffer does not contain potassium, we assume was bound to the protein. This is consistent with the requirement for potassium by the enzyme [11].

Amino acid sequencing and design of PCR primers

The N-terminus of ketohexokinase was not accessible to Edman degradation, and was therefore assumed to be blocked. The protein proved to be considerably resistant to enzymic cleavage. Combined treatment with trypsin (EC 3.4.21.4) and endoproteinase Glu-C (EC 3.4.21.19) eventually yielded peptides of a suitable size for amino acid sequencing. Fractions from the six largest peptide peaks resolved by reverse-phase h.p.l.c. were sequenced. One of these, KHK 43, was found to contain two peptides, yielding a pair of PTH-amino acid products for seven of the 21 cycles of Edman degradation. This ambiguity was later easily resolved from the cDNA, which allowed the distinction of peptides KHK 43a and KHK 43b.

Two peptides, KHK 45 and KHK 49, were found to have regions of sufficiently low degeneracy to allow the design and synthesis of small pools of oligonucleotides complementary to their amino acid sequences. Rather than employ these degenerate pools to screen a cDNA library directly, it was decided to use them as primers for the amplification of the region of cDNA that they flank. This strategy allows the generation of a PCR product with a substantial tract of DNA that is perfectly matched to the cDNA template. This fragment can therefore be used as a probe

to screen the cDNA library at very high stringency from the outset.

Synthesis and characterization of the PCR product

Total rat liver RNA was used as template for an oligo(dT)-primed synthesis of single-stranded cDNA. This was then employed as a template for PCR. Primers 2 and 3 (designed against peptides KHK 45 and 49) yielded a 224 bp PCR product. This was found to contain a section complementary to the amino acid sequence of a further peptide, KHK 22. The remaining sections of KHK 45 and 49 that were not represented in the oligonucleotide primers, but appeared in the PCR product, were consistent with the amino acid sequences obtained (Figure 3). We believed that these characteristics adequately confirmed the identity of the PCR product as being a fragment of the coding region of ketohexokinase cDNA, and proceeded to use it as a probe to screen the cDNA library.

Screening of the cDNA library

Twelve independent isolates, with inserts ranging in size from 1 kb to 2.5 kb, were obtained by plaque hybridization of approx. 10^6 clones. The three largest (2.3 kb, 2.4 kb and 2.5 kb) were rescued from the phage to yield the clones in the pBluescript SK(−) phagemid. Restriction mapping of the clones and preliminary sequencing of *Exo*III-deleted constructs of clone 16i revealed that all three contained substantial amounts of co-ligated, unassociated cDNA attached by distinctive *Eco*RI linkers (Figure 1).

Characterization and sequencing of clones

Both clones 16i and 5ii contained poly(A)$^+$ tails at their 3′-termini, which identified the cDNA as being a copy of fully or partially processed mRNA. The consensus polyadenylation signal AATAAA was identified 20 nucleotides upstream of each poly(A)$^+$ tract [32]. It was initially unclear whether clone 5ii contained the start codon and any associated 5′ untranslated region. Clone 5i contained the longest section of 5′ untranslated region, but did not contain the C-terminus of the coding region (Figure 1).

The hybrid clone 5i/16i (Figure 1) contained a section of cDNA encompassing the coding region and both the 3′ and 5′ untranslated regions, free of co-ligated cDNA. Sequencing revealed a point mutation in the parent clone 5i (AAT to GAT), resulting in a change of Asn-18 to Asp-18. Two lines of evidence supported the identification of this as a mutation. Firstly, the sequences of clone 5ii/a and the *Exo*III-deleted constructs of 16i encoded Asn-18; secondly, the amino acid sequence of peptide KHK 32 revealed an Asn at this position. The probability of this Asn to Asp change being a polymorphism is very slight, since the cDNA library and protein sequence were both obtained from the same strain of rat.

The complete nucleotide sequence and the corresponding predicted amino acid sequence of ketohexokinase are shown in Figure 3. The coding region has been verified using data obtained from both strands of at least two independent clones. The translation initiation codon was assigned using several criteria. Firstly, its position relative to the 3′ termination codon would generate a protein with a molecular mass (32728 Da) which is in close agreement with that determined by mass spectrometry (32800 Da). Secondly, all seven peptide sequences from the purified ketohexokinase lie in the open reading frame created by this start codon. Thirdly, it is the only in-frame ATG codon between the region encoding peptide KHK 32 and the in-frame

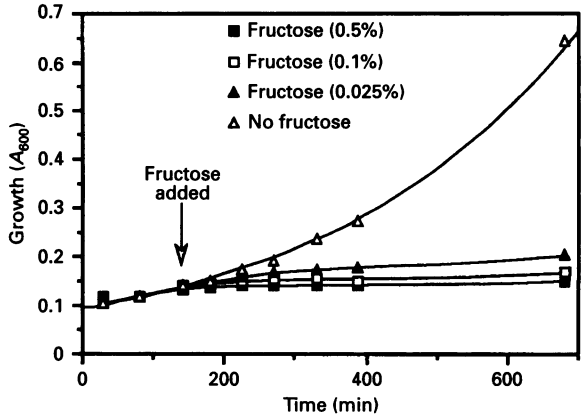

Figure 4 Growth profiles of the fructose-intolerant yeast strain TDB I upon addition of fructose at various concentrations

An exponential-phase culture of cells in SD Leu$^-$ medium was divided into equal aliquots, and the growth was monitored by measurement of A_{600}. After 2 h, fructose was added to the cultures to a final concentration of 0.5% (w/v) (■), 0.1% (w/v) (□) or 0.025% (w/v) (▲), or water was added as a control (△).

stop codon 198 nucleotides upstream of this initiation site. Fourthly, the overwhelming predominance of cytidine at the −1 and −4 positions of the initiation site sequences of rat cDNAs [33] is conserved within the ketohexokinase cDNA when it is assigned this start codon.

Expression of ketohexokinase in *Saccharomyces cerevisiae*

Fifteen transformants of the yeast TDS II bearing the plasmid pMA91-KHK (termed TDB I), and three bearing plasmid pMA91-KHK(−) [termed TDB I(−)], were inoculated on to SD Leu$^-$ medium. The presence of the ketohexokinase gene in transformants was confirmed by amplification of its coding region from plasmid DNA minipreps of these strains by PCR. Confirmation that the leucine-independence was borne episomally, and was not the result of reversion, was given by the loss of this phenotype after passage of the cells though non-selective YEPD medium.

Fructose-intolerance of TDB I strains was shown by their inability to grow when replica-plated on to SD Leu$^-$ medium containing 0.5% (w/v) fructose. TDB I(−) strains did not show intolerance when tested in this manner. The effects of various concentrations of fructose on the growth of a TDB I strain in liquid culture are shown in Figure 4. A similar culture of TDB I(−) was unaffected by the addition of fructose (results not shown). Metabolite extracts of TDB I cells exposed to 0.5% (w/v) fructose for 3 h showed the presence of fructose 1-phosphate (30 μmol of fructose 1-phosphate/g wet weight of cells). Fructose 1-phosphate was not present in TDB I(−) cell extracts.

Crude extracts from TDB I contained ketohexokinase with a specific activity of 0.62 μmol/min per mg of protein, which is almost nine times greater than that found in rat liver (0.07 μmol/min per mg of protein). Ketohexokinase activity was absent from extracts of both the TDS II and TDB I(−) strains. Control assays showed a minor ATPase activity in TDB I (0.069 μmol/min per mg of protein) which was absent from the parent TDS II strain. Curiously, in TDB I(−) this ATPase assumed a much higher activity (0.214 μmol/min per mg of protein).

Due to the co-migration of a large band of yeast protein in the crude extracts, partial purification of positive and control extracts

Table 1 Comparison of proteins using the MULTALIGN program

The percentage identity between the proteins is indicated above the diagonal, and the similarity score is shown below the diagonal. The similarity scores were determined by randomizing the sequence 100 times, and comparing each randomized sequence. The means and the S.D.s of the distribution of comparisons were calculated, and the similarity is expressed in terms of the number of units the S.D. is away from the mean of the random distribution. Abbreviations: KHK, ketohexokinase; pfk2, *E. coli* minor phosphofructokinase; FK, *V. alginolyticus* fructokinase; RK, *E. coli* ribokinase; GK and HK, rat liver glucokinase and hexokinase respectively.

	KHK	pfk2	FK	RK	GK	HK
KHK		21.4	18.24	19.66	20.27	21.21
pfk2	4.48		20.41	22.67	20.78	21.17
FK	3.24	6.17		23.21	16.99	19.61
RK	5.82	9.00	8.15		22.08	18.45
GK	0.03	−0.68	−0.33	−0.49		52.11
HK	1.24	−1.94	1.26	−0.50	59.23	

was necessary before a 31 kDa band corresponding to that for the ketohexokinase standard could be identified by SDS/PAGE.

Comparison of ketohexokinase sequence with other protein sequences

The National Biomedical Research Foundation (NBRF) Protein Identification Resource (PIR) V.29 databases were searched with the ketohexokinase protein sequence, which proved to be unique. Ketohexokinase does not contain sequence complementary to either the ATP- or hexose-binding sites that are common to members of the hexokinase family of proteins [7]. Pairwise comparison of amino acid sequences using the Alignment of Multiple Sequence (AMPS) program MULTALIGN [34] showed that no single protein had a prominently similar sequence identity with ketohexokinase, the greatest percentage identity being ~ 20 %. Of the protein sequences showing this degree of identity, three furanose sugar kinases were selected for detailed comparison, because of their similar size, function and substrate specificity. These proteins were the *E. coli* minor phosphofructokinase (EC 2.7.1.11) [35], *E. coli* ribokinase (EC 2.7.1.15) [36] and *Vibrio alginolyticus* fructokinase (EC 2.7.1.4) [37].

Multiple alignment of the ketohexokinase sequence with those of these three proteins was performed using the MULTALIGN program. This program uses the Needleman and Wunsch algorithm, modified to allow quick comparison of more than two sequences [34]. Rat hexokinase and glucokinase were also compared using this method, and the results are summarized in Table 1.

The multiple alignment of the four proteins revealed a highly conserved motif, TX(A/G)AGDX$_9$L, where the amino acids X are nearly always conservative amino acid replacements (Figure 5). This consensus was used to search the PIR databases and was also found in a fifth protein, *E. coli* DNA topoisomerase I [38].

DISCUSSION

The calculated molecular mass of rat liver ketohexokinase reported here (32728 Da) differs somewhat from the value obtained from the protein (28 kDa) by Sánchez et al. using density gradient centrifugation [13], but we are confident that our value is correct because it is in close agreement with that obtained by mass spectrometry. Although the protein has six potential glycosylation sites (Asn-Xaa-Ser/Thr), the mass spectrum showed no evidence of glycosylation.

The fructose-intolerance displayed by yeast expressing ketohexokinase is associated with a large, and essentially irreversible, accumulation of fructose 1-phosphate. The level of fructose 1-phosphate in TDB I cells treated with fructose was approximately three times the level of total sugar phosphates in yeasts performing aerobic glycolysis [39]. This accumulation of fructose 1-phosphate is similar to that observed in mammalian liver in subjects suffering from hereditary fructose intolerance [2]. In both cases, this is due to the absence of aldolase B activity when ketohexokinase remains active. Administration of fructose to rat

```
Ketohexokinase      MEEKQILCVG LVVLDIINVVDK YPEEDTDRRCLSQRWQR    GGNASNSCTVLSLLGARCAFMGSLAHGHVADFLVADFRRRGV
E. coli ribokinase  MQN     AGSLVVLGSINADHILNLQSFPTPGETVTGNHYQVAFGGKGANQAVAAGRSGANIAFIACTGDDSIGESVRQQLATDNI
E. coli pfk2         MVRIYTLTLAP SLDSATITPQIYPEENCAVPHRCSNP      GGGINVARAIAHLGGSATAIFPAGGAT GEHLVSLLADENV
Fructokinase              M NQVWVTGDAVVDLI PESETSLLKC         PGGAPANVAVAIARLSGKSAFFGRVGDDPFGRFMQSILDQEGV

Ketohexokinase      DVSQVAWQSQGDTPCSCCIVNNSNGSRTIILYDTNLPDVSAKDFEKVDLTRFKWIHIEGRNASEQVKMLQRIEQYNATQPLQQKVRV
E. coli ribokinase  DITPVSVIKGESTGV ALIFVNGEGENVIG  IHAGANAALSPA LVEAQRERIANASALLMQLESPLESVMAAAKIAHQNKTIVAL
E. coli pfk2        PVATVE  AKDWTRQNLHVHVEASGEQ YRF VMPG  AALNEDEFRQLEEQVLEIESGAILVISGSLPPGVKLEKLTQLISLRKNK
Fructokinase        CTEFLIKDPEQRTSTVVVDLDD QGERSFTFMVKPSADQFMSVEDMGNFKQGDWLHVCSISLANEPSRSSTFEAIKRAKAAGGFISF

Ketohexokinase      SVEIEKPR EELFQLFGYGEVVFVSKDVAKHLGFR     SAGEALKGLYSRVKKGATLICAWAEEGADALGPDGQLLHSDAFPP
E. coli ribokinase  NPAPARELP DELLALVDIITPNETEAE KLTGIRVEND   EDAAKAAQVLHEKGIRT VLITLGSRGVWASVNGEGQRVPGFR
E. coli pfk2        GSAASSTVLGQGLSAALAIGNIELVKPNQKELSALVNRELTQPDDVRKAAQEIVNSGKAKRVVVSLGPQGALGVDSEN CIQVVPPA
Fructokinase        DPNLRDEVW QDQSEIQAVVMKAVAMADVVKFSEEELLFLTDSTSMAQGLQQIAAMNIAL VLVTQGAKGVWRVFESQSELITGQV

Ketohexokinase      PKVVDTIGAGDTFNASVIFSLS      KGNSMQEALRFGCQVAGKKCGLQG          FDGIV
E. coli ribokinase  VQAVDTIAAGDTFNGALITALL      EEKPLPEAIRFAHAAAAIAVTRKGAQPSVPWREE IDAFLDRQR
E. coli pfk2        LKSQSTVGAGDRLVGAMTLKLA      ENASLEEMVRFGVAAGSAATLNQGTRLCSHDDTQKIYAYLSR
Fructokinase        VSPIDTTGAGDAFVGGLLACLSRHADWKNHPVVSSAIQWANGCGALATTQKGAMTALPTQTE LLRFIGQ
DNA topoisomerase   DSTTLTVGAGDFRLKARGRIIR
```

Figure 5 Comparison of amino acid sequences of ketohexokinase and other furanose kinases

The sequences were compared using the MULTALIGN program. The amino acid similarities of ribokinase, phosphofructokinase 2 (pfk2) and fructokinase are compared with those of ketohexokinase. Those found to be identical are in bold, and those with conservative replacements are italicized. The boxed sections contain a region of high-sequence identity, the consensus of which was used to search the PIR databases. This led to the identification of DNA topoisomerase I, the relevant sequence of which is indicated below these sections (amino acid residues 414–435). The sequences of the four proteins were taken from the NBRF PIR (v. 29) database, and have the following accession numbers; ribokinase, KIECRB; pfk2, KIECFB; *Vibrio alginolyticus* fructokinase, JQ0782; DNA topoisomerase I, ISECTP.

liver results in a transient depletion of ATP and cytoplasmic orthophosphate in the tissue owing to the sequestration of phosphate in fructose 1-phosphate. The absence of aldolase B in patients suffering from hereditary fructose intolerance will prevent recovery from such depletions. This causes a severe compromise of energy metabolism, although an exact locus at which metabolic paralysis is imposed has not been identified. It is clear that the fructose-intolerant yeast (TDB I) suffers equally severe metabolic impairment on exposure to fructose, and could serve as a metabolic model to study this type of disorder.

It is our intention to exploit the fructose-intolerance of yeast expressing ketohexokinase to select for mutations of invertase that retain fructotransferase activity [40], but will not hydrolyse sucrose. The genotype of the TDS II strain renders it incapable of utilizing sucrose as a carbon source. A fructose-intolerant strain derived from this will only survive on sucrose as its sole carbon source if it is expressing a mutated invertase that sequesters fructose as a fructofuranoside and only releases glucose. Under these circumstances, cells expressing wild-type hydrolase activity will die due to the release of fructose, and cells expressing inactive mutations of invertase will not grow due to the lack of a utilizable carbon source.

Interestingly, the amino acid sequence of ketohexokinase reveals that it is much more closely related to the prokaryotic furanose kinases examined than it is to the eukaryotic hexokinase family. Table 1 predictably shows a high percentage identity when a direct comparison is made between glucokinase and hexokinase. The percentage identities obtained with all other pairwise comparisons are $\sim 20\,\%$. However, the similarity scores clearly show that, although glucokinase and hexokinase are very closely related to each other, they show no similarity with the other proteins. The similarity scores obtained for all possible comparisons between the other proteins show that they are related to one another, although, unsurprisingly, the mammalian ketohexokinase appears to be the most evolutionarily divergent protein within the group. The most striking similarity between the proteins is the conserved region depicted in Figure 5. Although these proteins are all ATP-dependent kinases, the possibility that this region encodes a common binding site for ATP would seem to be discounted by its conservation in the structural gene for DNA topoisomerase I, which does not employ ATP or NAD^+ as a substrate. The proposed mechanism for *E. coli* topoisomerase I [41] involves the transient breakage of a DNA backbone bond and simultaneous formation of a phosphotyrosine bond between the protein and the cut 5′ phosphoryl group. The cleavage specificity suggests that the enzyme is interacting with both sides of the point it cleaves, and therefore bridges the breakpoint, binding non-covalently with the 3′ deoxyribose sugar. The conserved motif $TX(A/G)AGDX_9L$ may be involved in the non-covalent binding of either the 5′ deoxyribose phosphate group or the 3′ sugar. The conservation of this region may therefore reflect a common function in the binding of furanose sugars.

Ketohexokinase has recently been reported in the halophilic archaebacterium (*Haloarcula vallismortis*) [42]. This is the first report of an ATP-dependent D-fructose 1-phosphotransferase in prokaryotes. It would be interesting if future studies could show whether this protein belongs within the same group of furanose kinases revealed in this paper, as this would lend support to our belief that this represents an ancient family of proteins.

Received 7 September 1992/29 October 1992; accepted 10 November 1992

We are indebted to Dr. David Bonthron of the University of Edinburgh for his technical advice and for critically reading this manuscript. We also thank Dr. David Harvey for assistance with the mass spectrometry, Mr. Tony Willis for performing the amino acid sequence analysis, and Dr. Geoff Barton for his assistance with the sequence comparisons. This work was supported by the University of Oxford. T.C.D. was in receipt of a Science and Engineering Research Council studentship. N.M. was supported by scholarships from the University of Oxford and the British Council.

REFERENCES

1 Cori, G. T., Ochoa, S., Slein, M. W. and Cori, C. F. (1951) Biochim. Biophys. Acta **7**, 304–317
2 Van den Berghe, G. (1986) Prog. Biochem. Pharmacol. **21**, 1–32
3 Malaisse, W. J., Malaisse-Lagae, F., Davies, D. R. and Van Schaftingen, E. (1989) FEBS Lett. **255**, 175–178
4 Cross, N. C. P., Tolan, D. R. and Cox, T. M. (1988) Cell **52**, 881–885
5 Schapira, F., Schapira, G. and Dreyfus, J. C. (1961/1962) Enzym. Biol. Clin. **1**, 170–175
6 Càrdenas, M. L., Rabajille, E. and Niemeyer, H. (1984) Biochem. J. **222**, 363–370
7 Griffin, L. D., Gelb, B. D., Wheeler, D. A., Davison, D., Adams, V. and McCabe, E. R. B. (1991) Genomics **11**, 1014–1024
8 Walker, D. G. (1968) Essays Biochem. **2**, 33–67
9 Raushel, F. M. and Cleland, W. W. (1973) J. Biol. Chem. **248**, 8174–8177
10 Raushel, F. M. and Cleland, W. W. (1977) Biochemistry **16**, 2169–2175
11 Sánchez, J. J., González, N. S. and Pontis, H. G. (1971) Biochim. Biophys. Acta **227**, 79–85
12 Adelman, R. C., Ballard, F. J. and Weinhouse, S. (1967) J. Biol. Chem. **242**, 3360–3365
13 Sánchez, J. J., González, N. S. and Pontis, H. G. (1971) Biochim. Biophys. Acta **227**, 67–78
14 Bais, R., James, H. M., Rofe, A. M. and Conyers, R. A. J. (1985) Biochem. J. **230**, 53–60
15 Romanos, M. A., Scorer, C. A. and Clare, J. J. (1992) Yeast **8**, 423–488
16 Bensadoun, A. and Weinstein, D. (1976) Anal. Biochem. **70**, 241–250
17 Hames, B. D. (1981) in Gel Electrophoresis of Proteins (Hames, B. D. and Rickwood, D., eds.), pp. 1–86, IRL Press, Oxford
18 Beavis, R. C. and Chait, B. T. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 6873–6877
19 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
20 Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. **162**, 156–159
21 Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene **33**, 103–119
22 Bonthron, D. T. (1990) Gene **91**, 101–105
23 Sanger, F. (1981) Science **214**, 1205–1210
24 Short, J. M., Fernandez, J. M., Sorge, J. A. and Huse, W. D. (1988) Nucleic Acids Res. **16**, 7583–7600
25 Friedman, K. D., Rosen, N. L., Newman, P. J. and Montgomery, R. R. (1990) in PCR Protocols: A Guide to Methods and Applications (Innis, M. A., Gelfand, D. H., Sninsky, J. J. and White, T. J., eds.), pp. 253–257, Academic Press, San Diego
26 Hamilton, R., Watanabe, C. K. and de Boe, H. A. (1987) Nucleic Acids Res. **15**, 3581–3593
27 Mellor, J., Dobson, M. J., Roberts, N. A., Tuite, M. F., Emtage, J. S., White, S., Lowe, P. A., Patel, T., Kingsman, A. J. and Kingsman, S. M. (1983) Gene **24**, 1–14
28 Ito, H., Fukuda, Y., Murata, K. and Kimura, A. (1983) J. Bacteriol. **153**, 163–168
29 den Hollander, J. A., Ugurbil, K. and Shulman, R. G. (1986) Biochemistry **25**, 212–219
30 Eggleston, L. V. (1983) in Methods of Enzymatic Analysis (Bergmeyer, H. U., ed.), vol. 2, pp. 328–334, Verlag Chemie, Weinheim
31 Jazwinski, S. M. (1990) Methods Enzymol. **182**, 154–174
32 Proudfoot, N. and Brownlee, G. G. (1976) Nature (London) **263**, 211–214
33 Kozak, M. (1984) Nucleic Acids Res. **14**, 857–872
34 Barton, G. J. (1990) Methods Enzymol. **183**, 403–428
35 Daldal, F. (1984) Gene **28**, 337–342
36 Hope, J. N., Bell, A. W., Hermodson, M. A. and Groarke, J. M. (1986) J. Biol. Chem. **261**, 7663–7668
37 Blatch, G. L., Scholle, R. R. and Woods, D. R. (1990) Gene **95**, 17–23
38 Tse-Dihn, Y.-C. and Wang, J. C. (1986) J. Biol. Chem. **261**, 321–331
39 Brindle, K. M. (1988) Biochemistry **27**, 6187–6196
40 Edelman, J. (1956) Adv. Enzymol. **17**, 189–232
41 Wang, J. C. (1985) Annu. Rev. Biochem. **54**, 665–697
42 Altekar, W. and Rangaswamy, V. (1991) FEMS Microbiol. Lett. **83**, 241–246