Cloning and characterization of the DAS gene encoding the major methanol assimilatory enzyme from the methylotrophic yeast Hansenula polymorpha

Zbigniew A. Janowicz, Michael R. Eckart, Christel Drewke, Rainer O. Roggenkamp and Cornelis P. Hollenberg  
Institut für Mikrobiologie, Universität Düsseldorf, 4000 Düsseldorf, FRG, and

Jan Maat, Aat M. Ledeboer, Chris Visser and C. Theo Verrips  
Unilever Research Laboratorium Vlaardingen, P.O. Box 114, 3130 AC Vlaardingen, The Netherlands

Received 4 March 1985; Accepted 10 April 1985

ABSTRACT  
A gene library from the methanol utilizing yeast Hansenula polymorpha, constructed in a λ Charon4A vector, was used to clone the gene encoding a key methanol assimilating enzyme, dihydroxyacetone synthase (DHAS) by differential plaque hybridization. The nucleotide sequence of the 2106 bp structural gene and the 5' and 3' non-coding regions was determined. The deduced amino acid sequence of the protein is in agreement with the apparent molecular weight and amino acid composition of the purified protein. The codon bias is not so pronounced as in some Saccharomyces cerevisiae genes.

INTRODUCTION  
The ability of certain yeast strains to utilize methanol as a sole carbon- and energy source is currently receiving increasing attention from both basic research and biotechnology (1-3). Research has however mainly been restricted to enzymatic studies of the methanol dissimilatory pathway. Two methylotrophic yeasts, the asporogenous species Candida boidinii and the thermotolerant ascomycete Hansenula polymorpha have been used to study the above aspects in detail (1-3). In both species methanol is oxidized via formaldehyde and formate to CO₂ by methanol oxidase, MOX, (1,4) and two dehydrogenases (1,5). Catalase serves to remove H₂O₂, which is formed in the first oxidation step (3).

The key reaction in the C₁-compound assimilation is the transfer of a glycoaldehyde group of xylulose 5'-phosphate to formaldehyde, generated in the dissimilatory pathway, resulting in the formation of dihydroxyacetone and glyceraldehyde-3-phosphate. This specific transketolase was given the trivial name dihydroxyacetone synthase, DHAS (6,7).

Growth of H. polymorpha on methanol is accompanied by dramatic
changes in cell protein composition (8,9) and the appearance of specific organelles, peroxisomes, bearing methanol oxidase and catalase (3,10). Thus, H. polymorpha is a good model system to study the function, biogenesis and turnover of peroxisomes. H. polymorpha methanol assimilatory and dissimilatory enzymes are induced by methanol and subjected to glucose repression (2,9).

In methanol grown cells the key enzymes of methanol utilization, DHAS and MOX, constitute approximately 15 and 20% of total cell protein, respectively (1-4,9). Previously, we have shown that the mRNAs encoding these proteins are predominant transcription products indicating the existence of very efficient promoters of the respective genes (9).

We aimed at cloning and characterising the genes involved in methanol utilization to obtain data at the molecular level of the regulation of this complex and highly compartmentalized metabolic system. Expression studies of these genes in homologous and heterologous systems should bring more insight into the regulatory elements involved in lower eukaryotes, since most of the data available derive from the model organism Saccharomyces cerevisiae (11). Also, a close characterization of gene regulation of the key enzymes in a methylotrophic yeast could prove to be of considerable biotechnological importance. In this paper we report the cloning of the gene coding for the assimilatory enzyme DHAS, from a Charon4A bank of H. polymorpha. In order to be able to perform further regulation studies we have sequenced the 2106 bp of the coding region and its 5'- and 3'end flanking regions. The 5'regulatory region shares only a few common features with S. cerevisiae genes whereas the 3'end region shows a remarkable resemblance with them. The codon usage in this gene differs from sequenced S. cerevisiae genes. In the accompanying paper the cloning and the analysis of the dissimilatory MOX gene is presented (12).

MATERIAL AND METHODS
Strains, media and vectors. A thermophylic, homothallic strain of H. polymorpha (ATCC 34438) was used as a wild type strain. Yeast was grown on minimal YNB media containing 0.67 % yeast nitrogen base without amino acids (Difco), 0.1 % yeast extract (Difco) and 0.05 % peptone (Difco). The minimal media were supplemented with
3% glucose to achieve a full repression of enzymes involved in methanol utilization, or with 1% methanol for induced growth conditions. Cells were grown at 30°C with vigorous shaking. *Escherichia coli* K12 strains used were: *E. coli* 490; C600 recA, hsdM, araB used for transformation with recombinant plasmids; JM103, thi, strA, supE, endA sbcB, hsdR, F'traD36, proAB, lacI, ZM15 as a host for propagation of phage M13; KH802 gal, met, supE host for λ vector Charon4A. *E. coli* strains were grown in L-broth media without glucose. Plasmid YRp7, consisting of pBR322, a yeast *S. cerevisiae* autonomous origin of replication (ARS1) and the yeast TRP1 gene (13) was used for subcloning of *H. polymorpha* DNA.

Plasmid DNA and DNA from the replicative form of M13 recombinant phages were isolated by a scaled up alkaline minilysate method (14) followed by CsCl/ethidium bromide ultracentrifugation. Charon4A (15) was used to construct a *H. polymorpha* DNA library. DNA from Charon4A and its recombinant clones was isolated by a scaled up plate lysate method (16).

Restriction enzymes, T4 ligase, Klenow polymerase, polynucleotide kinase were purchased from Boehringer, Mannheim. The conditions for the enzymatic reactions were as recommended by the supplier unless otherwise stated.

**Isolation of DNA from *H. polymorpha* spheroplasts.** Cells were harvested at early log-phase after growth in 3% glucose minimal medium, washed twice with H2O and then incubated at 30°C for 10 min in 0.9 M sorbitol, 50 mM Tris, 50 mM EDTA, 48 mM β-mercaptoethanol (final pH 9.0), washed with 0.9 M sorbitol and incubated at 30°C with Zymolyase 5000 (Kirin), 10 μg/ml, in 0.9 M sorbitol, 50 mM Na-phosphate, 50 mM EDTA (pH 7.4) in a volume of 100 ml per 1 g of cells. Spheroplasts were then gently washed twice in 0.9 M sorbitol, 0.1 M EDTA (pH 7.4). The spheroplast pellet was lysed in 0.1 M EDTA (pH 8.5), 50 mM NaCl, 50 mM Tris (pH 8.5), 1% sarcosyl (100 ml lytic mixture per 5g of cells). After 10 min the lysate was gently extracted twice with phenol. To the aqueous phase CsCl was added (1.3 g/ml) and the DNA was purified by ultracentrifugation in the absence of ethidium bromide (44,000 rpm, Vti50 vertical Beckmann rotor, 48 h). In a typical experiment 1 g of cells yielded 300 μg of DNA of a size greater than 50 kb.

Charon4 *H. polymorpha* library. *H. polymorpha* DNA was partial-
ly digested with EcoRI (0.2 units/µg DNA) for 20 min at 37°C, yielding partial fragments of a size of 10 - 30 kb. The fragments were fractionated by sucrose density gradient ultracentrifugation to obtain a 20 kb fraction (17). These fragments were ligated to sucrose gradient purified (17) Charon4 arms. The ligated DNA was converted into infectious phage particles by in vitro packaging (18), yielding 300,000 pfu per µg of ligated DNA. E. coli lyso- gens BHB2690 (Dam) and BHB2688 (Eam) were used to prepare packaging extracts.

PolyA mRNA from H. polymorpha was prepared as described (9) with a few modifications. Fractionation of polyA mRNA on a 10-30% sucrose gradient was performed as described previously (9). Some mRNA fractions, enriched in mRNA species encoding MOX and DHAS, were additionally purified by disc polyacrylamide (3.1 %) gel electrophoresis, a system described for isolation of rRNA (19). The procedure yielded mRNA fractions which contained about 70–80% of MOX and DHAS mRNA (see Results Fig.1, lane 5).

Analysis of mRNA by an in vitro cell-free rabbit reticulocyte system, and subsequent immunoprecipitation of translation products was performed as described previously (9). Rabbit antisera against MOX and DHAS proteins, purified to homogeneity, were prepared as described in our previous publication (9).

32P-labeled cDNA was synthesized from polyA mRNA using Avian Myeloblastosis virus reverse transcriptase. The enzyme (Beard enzyme) and other reagents were purchased from New England Nuclear. The length of the cDNA transcripts ranged from 600-900 bp, and their specific activity was 2 x 10^7 cpm per µg of mRNA used (NEN 32P-dCTP, 400 Ci/mmol was used). RNA was 5'end labeled with 32P-ATP (Amersham) after fragmentation by mild alkaline treatment as described (20).

The differential plaque filter hybridization was performed as described (21,22).

Sequence Analysis. Different DNA fragments, isolated from low melting agarose gels, were cloned into the phage M13 derivatives mp8 and mp9 (23) and sequenced by the dideoxy chain termination method of Sanger et al. (24,25). Small inserts cloned in two orientations were sequenced directly from both sides, whereas longer fragments were sequenced using the Bal31 exonuclease digestion strategy (26), see also accompanying paper (12).
Si-mapping experiments were performed according to Favarolo et al. (27). The hybridization of DNA fragments with mRNA was performed in 40 mM PIPES (pH 6.4), 1 mM EDTA, 0.4 M NaCl, 80% formamide at 45°C, 48°C or 50°C for 3-6 hr. Si nuclease digestion (Boehringer) was carried out with 1000 units/ml at 30°C or 37°C for 1 h.

Selection of specific mRNA species (hybrid selection technique) was achieved by hybridization of mRNA to subcloned DNA which was covalently bound to diazophenylthioether (DPTE) derivative of Sephacryl S-500 exactly as described by Bünemann et al. (28,29). Modified Sephacryl S-500 matrix was a gift from Dr. Bünemann, University of Düsseldorf. mRNA selected by this procedure was then translated in an in vitro cell-free reticulocyte translation system, and the translation products analysed by immunoprecipitation with specific antibodies as described in our previous paper (9). In a typical experiment 50 µg DNA was immobilized on the DPTE-S-500 matrix and hybridized with 25 - 50 µg of polyA mRNA.

Peptide mapping by limited proteolysis in SDS and subsequent analysis by SDS-PAGE was performed exactly as described by Cleveland et al. (30). Protease V8 was used in these experiments.

RESULTS
Identification of mRNA species encoding DHAS and MOX. The total polyA mRNA isolated from mid-log phase cultures of H. polymorpha grown on either 3% glucose (not induced mRNA) or 1% methanol (induced mRNA), was fractionated by sucrose density gradient centrifugation. Some of these fractions were additionally purified by polyacrylamide gel electrophoresis (PAGE)(see Material and Methods and Fig.1, lane 5). Both fractionated and unfractionated mRNA preparations were translated in an in vitro cell-free reticulocyte translation system and the translation products analysed by SDS-PAGE (Fig.1). As reported previously (9) a comparison of not induced (Fig.1, lane 7) and induced (lane 6) mRNA in vitro translation products with the in vivo products (lanes 1-4) enabled us to identify the MOX and DHAS mRNA species. Fig.1 clearly shows that the MOX and DHAS proteins (lane 4) and their respective mRNA species (lane 6) are predominant in cells grown on methanol.
Fig. 1: Analysis and identification of MOX and DHAS proteins and their respective mRNA by an 8-15% gradient SDS polyacrylamide gel electrophoresis.

Lanes 1-4. Coomassie Blue stained gel. 1, purified dihydroacetone synthase (DHAS); 2, purified methanol oxidase (MOX); 3, crude extract of cells grown under repressed conditions (3% glucose); 4, crude extract of induced cells (1% methanol).

Lanes 5-12. Fluorography of \(^{35}\)S Methionine labeled in vitro translation products of different mRNA. 5, purified mRNA enriched in DHAS and MOX mRNA species; 6 and 7, translation products of mRNA from induced and not induced cells, respectively; 8, products of hybrid selected mRNA complementary to DNA from clone 7/7; 9-11, translation products from lane 8 precipitated with anti DHAS serum, anti MOX serum and preimmune serum, respectively; 12-13 immunoprecipitation of translation products from total induced polyA mRNA with anti DHAS serum (lane 12) and anti MOX serum (lane 13).

Screening for genes involved in methanol utilisation. \(^{32}\)P-labeled cDNA as well as \(^{32}\)P-5'end labeled mRNA (see Material and Methods) were prepared from the following polyA mRNA preparations: (i) unfractionated induced mRNA, (ii) unfractionated not induced mRNA, and (iii) fractionated induced mRNA enriched in mRNAs encoding MOX and DHAS (Fig. 1, lane 5). Additionally, \(^{32}\)P-5' end labeled rRNA was prepared as a hybridization probe to eliminate the clones containing ribosomal DNA sequences. A set of 6 or 7 replica nitrocellulose filters was prepared from each plate containing 3000 recombinant phages and screened by individual hybridization to the radioactive probes described above. One set of filters was hybridized in a solution containing \(^{32}\)P-cDNA from en-
Identification of the mRNA species encoded by DNA from the recombinant clone 7/7 using Northern blotting. Lanes 1 and 2, hybridization of radioactive cDNA prepared from induced total mRNA to not induced mRNA (1) and induced mRNA (2). Lanes 3-5, hybridization of the radioactive recombinant clone 7/7 to the following total polyA mRNA preparations: 3, not induced mRNA (10 µg); 4, not induced mRNA (100 µg); 5, induced mRNA (10 µg). Lane 6, hybridization of the 4.5 kb BamHI clone, encompassing the DAS gene, to induced mRNA (5 µg).

Identification of isolated recombinant clones. We aimed at selecting clones which hybridize only to induced mRNA of the size able to encode the MOX or DHAS protein. Fig.2 shows a Northern blot (31) obtained by hybridization of radioactive recombinant clone 7/7 to induced (lane 5) and not induced mRNA (lanes 3,4). Additionally, 32P-cDNA synthesized from induced mRNA was hybridized to induced (lane 2) and not induced (lane 1) mRNA. A comparison of lanes 2 and 5 indicates that clone 7/7 contains a DNA sequence coding for a predominant induced mRNA species. No hybridization of clone 7/7 to not induced mRNA was observed even when a 10-fold excess over induced mRNA was present on the blot (lanes 3 and 4). The size of the hybridizing mRNA (about 2.3 kb) allows...
the assumption that it codes for a protein of the Mr of approximately 70,000-78,000. This fact helped to eliminate the possibility that the selected clone might code for one of the other strongly or weakly inducible proteins already reported (1,8,9) such as catalase or formate dehydrogenase since all of these exhibit an apparent Mr far below the coding capacity of this mRNA species. The above presented data, however, could not distinguish as to whether clone 7/7 might code for DHAS or MOX.

To distinguish between these two possibilities, we have isolated the mRNA species complementary to clone 7/7 by hybrid selection and analysed its translation product. Total polyA mRNA from methanol-grown cells was hybridized to the immobilized clone 7/7 DNA. The mRNA species not complementary to the DNA were washed out under stringent conditions. Hybrid selected mRNA eluted from the DNA-matrix was then translated in a cell-free rabbit reticulocyte system and its radioactive translation products compared with those of total induced and not induced polyA mRNA. Fig.1 shows that the hybrid selected mRNA yields two polypeptide bands (lane 8), the upper of the two having the same apparent Mr as the purified DHAS (lane 1), the other obviously a bit smaller than the MOX polypeptide. The same pattern was obtained when hybrid selected mRNA complementary to a 1260 bp long EcoRI/BamHI fragment containing the 3'end part of the gene in question was used in the experiment (see next chapter and Fig.4 for localisation of the gene fragments on clone 7/7). Immunoprecipitation of translation products of hybrid selected mRNA revealed that specific antibodies raised against the DHAS protein precipitate both bands (lane 9) whereas MOX antiserum and a preimmune serum do not show any cross-reactivity (lanes 10,11). Moreover when the in vitro translation products of total polyA induced mRNA were precipitated with DHAS antibodies an identical pattern of two bands appeared (lane 12). In the same experiment a preimmune serum gave no detectable precipitation and MOX antiserum gave one band at the position of purified MOX protein (lane 13). When, however, the in vivo synthesized proteins of methanol induced cells were analysed on high resolution gels, we found no indication of a double band for DHAS (lane 4). The same holds true for purified DHAS protein (lane 1).
Fig. 3: A Cleveland peptide mapping of the in vitro translation products obtained from hybrid selected mRNA (Fig. 1, lane 8) and from immunoprecipitation of induced mRNA translation products with anti DHAS serum (Fig. 1, lane 9) was performed. The two bands which resulted in both instances were separated on a high resolution 7-10 % polyacrylamide gel, isolated from the gel, digested in situ with 12.5 μg of V8 protease and the resulting peptides separated on a 12-20 % gradient polyacrylamide gel. Lanes a and b compare the polypeptide pattern of the upper (lane a) and lower (lane b) band obtained after in vitro translation of hybrid selected mRNA. In lanes c and d, the polypeptide pattern of the upper and lower bands obtained after immunoprecipitation of translation products of total induced mRNA are compared.

To check as to whether the two bands in question contain the same material, as suggested by the immunoassays, we have used the Cleveland peptide mapping technique (30). The two 35S-labeled bands derived from in vitro translated hybrid selected mRNA were excised from a high resolution gel and their polypeptide patterns were compared after in situ partial digestion with V8 protease. A series of mapping experiments with varying amounts of V8 protease revealed that within the limits of this technique no significant differences between the two bands can be observed (Fig. 3, lanes a,b). The same holds true for material obtained by immunoprecipitation of translation products of total induced polyA mRNA (Fig. 3, lanes c,d). Fig. 3 shows an example of the mapping experiments and indicates that a hybrid selected mRNA and total induced polyA mRNA give upon in vitro translation the same products.

The data presented in Fig. 1 and 3 strongly suggest that the cloned gene is encoding the DHAS protein. The gene was designated DAS.
Fig. 4: (A) Restriction map of clone 7/7 containing the DAS gene and the 4.5 kb BamHI fragment subcloned into the vector YRp7. The cleavage sites shown were determined by restriction endonuclease digestion and electrophoresis of generated fragments through agarose gels. Positions were confirmed by DNA sequencing. (B) The arrow indicates the direction of transcription as determined by S1 nuclease mapping experiments. The various fragments used for S1 mapping experiments are as indicated. (C) Sequence strategy of the subcloned 4.5 kb BamHI fragment. The indicated M13 subclones were sequenced up to 300 bp from their 5' ends. M13 HindIII-EcoRI subclones were cleaved with SmaI, subsequently treated with Bal31 nuclease and the samples taken during digestion were religated giving rise to new subclones which were also sequenced up to 300 bp from their 5' ends. The bars indicate the sequenced regions in coding (upper) and non-coding (lower) strands. Additional sequence data was obtained by ultrasonic shearing of the HindIII-SmaI and SmaI-BamHI fragments and the subsequent cloning of the generated fragments in a suitable M13 vector. In case of uncertainties specific primers were synthesized to sequence these regions.

Restriction mapping and subcloning of the DAS gene. Clone 7/7 was subjected to restriction enzyme analysis yielding the map of restriction sites indicated in Fig. 4A. Localisation of the DNA fragment encoding the DHAS protein was achieved by Southern hybridization (32). As a hybridization probe 32P-5'end labeled fractionated induced mRNA was used. Analysis of the hybridization patterns enabled us to localize the gene on a single 4.5 kb long
BamHI-BamHI fragment (Fig. 4A). This fragment was subcloned into the plasmid vector YRp7 which contains an E. coli and a S. cerevisiae origin of replication, enabling it to replicate in both organisms (13). The plasmid containing the gene was named YRp7-DAS. By Northern hybridization, it could be shown that the subcloned BamHI fragment and the original recombinant clone 7/7 hybridize to the same RNA band (Fig. 2, lanes 5 and 6).

S1-nuclease mapping. Size and direction of transcription of the DAS gene. The 4.5 kb BamHI fragment was cut at its single EcoRI site and a 3'end radioactive label introduced at the two resulting EcoRI ends. From the two labeled EcoRI/BamHI fragments, 3190 bp and 1260 bp (Fig. 4B), only the label on the right 1260 bp fragment was found to be protected against the action of S1 nuclease by hybridization with induced mRNA. The determination of the exact 3'end of mRNA was complicated by the fact that in the region following the open reading frame (see further below) the GC content is 37% whereas in the reading frame this value is 51%. This caused the values obtained from S1 mapping to be strongly dependent on the hybridization temperature. When this temperature was 45°C the S1 mapping yielded two bands, a weak one of about 1070 bp and a strong one of about 1120 bp (Fig. 5, lane 1). At 50°C the 3'end region hybrid was less stable giving a reverse situation, strong 1070 bp band and weaker 1120 bp band (Fig. 5, lane 2). This result indicated that the direction of transcription of the DAS gene is from left to right in Fig. 4, and that the 3'end of transcription is located about 180 - 140 bp prior to the right BamHI site of the 4.5 kb fragment. Since the Mr of the DHAS protein is about 78,000, the 4.5 kb BamHI fragment should contain the whole coding region together with regulatory elements.

Different labeled or unlabeled restriction fragments derived from the 4.5 kb subclone (Fig. 4B) were used for further S1 mapping studies. DNA fragments protected by mRNA were analysed on PAGE/urea or and on denaturing agarose gels. When non-radioactive fragments were used the bands were visualized by Southern blotting and subsequent hybridization with appropriate radioactive probes. The 2.3 kb BamHI/SmaI fragment containing a 5'-label at its SmaI site yielded a broad range of fragments from 135-155 bp long (lane 5). This confirmed the already established dire
Fig. 5: Identification of transcriptional start- and endpoints of the DAS gene by S1 mapping. S1-digestion products were separated on alkaline 1.2% agarose (lanes 1-4 and lanes 8-10) or on 5% polyacrylamide containing 8 M urea (lanes 5-7). M lanes are Mr markers. The fragments used for the mapping are described in Results and shown in Fig. 4B. 1260 bp EcoRI/BamHI fragment with 3' end label at EcoRI site was hybridized with the following mRNAs: 25 µg induced mRNA at 45°C (lane 1), and at 50°C (lane 2), tRNA without S1 treatment (lane 3), tRNA with S1 treatment. 2200 bp BamHI/SmaI fragment 5' labeled at SmaI site hybridized with: induced mRNA at 48°C (lane 5), tRNA without (lane 6) and with S1 treatment (lane 7). Lanes 8,9 - protection of the whole gene by its cognate mRNA; an unlabeled 4.5 kb BamHI fragment is shortened to 2.2 kb (lane 8) by S1 treatment. Lane 9 - mRNA protects a 2.2 kb fragment when recombinant clone 7/7 was used for S1 mapping. Lane 10 - 4.5 kb fragment without S1 treatment. Hybridization temperature was 48°C. When unlabeled fragments were used, the bands were visualized by Southern blotting followed by hybridization with the appropriate radioactive probe.

The identification of transcription and indicated multiple transcription starts about 135 - 155 bp left of the single SmaI site as indicated in Fig. 4. Comparison of this data with sequence analysis (see next chapter) allowed us to locate these starts approximately at positions -20 to -35 from the ATG initiation codon. Lane 8, Fig. 5, shows a continuous protection of about 2200 bp when the whole 4.5 kb BamHI fragment was subjected to S1-mapping. The same result was obtained when the whole clone 7/7 was used for the experiment (lane 9). This indicates that the gene does not contain a
detectable intron region. Remarkable is the fact, that not induced mRNA, even in excess (100 μg polyA mRNA per 2 μg DNA), did not protect any part of the gene region.

The protection of DNA by its cognate mRNA against S1 digestion was also used to estimate the amount of DAS mRNA in total polyA mRNA isolated from methanol grown cells. 10 μg of 3′ end-labeled 1260 bp EcoRI/BamHI fragment was hybridized with different amounts (10-50 μg) of polyA mRNA and digested with S1 nuclease. The resulting labeled DNA fragment was separated on an agarose gel and the radioactivity in the labeled band was measured. At the DNA excess conditions, 10 μg of induced polyA mRNA was protecting 0.7 μg of single stranded DNA which represents the level of DAS mRNA being about 7% of a total polyA mRNA preparation isolated at mid log phase.

Nucleotid sequence of DAS and its flanking region. To be able to study the transcriptional regulatory elements of the gene and to analyse the protein structure, we sequenced the 3.2 kb HindIII/BamHI fragment encompassing the gene. Fig.4C shows the sequencing strategy indicating that for most of the gene regions two strands were sequenced. Fig.6 shows the sequence of the structural gene and the 5′ and 3′ regulatory regions. Prior to the open reading frame (702 amino acids) several stop codons are present in all reading frames. Within 21 bp from the translation stop codon (TAA) there are two additional ones (TAA). The deduced amino acid composition indicates a protein of Mr of 77,000 which fits to the apparent Mr calculated for DHAS from SDS-PAGE (9). The GC content in the sequenced gene is 51%. The presumptive control region around the transcriptional starts shows several remarkable features, some of them being similar to corresponding regions of highly expressable yeast genes. Usually, purines have been found at positions -3 and +4 (33) and a pyrimidine at position +6 (11). In the H. polymorpha DAS gene a pyrimidine (T) is present at position +6, a purine at +4 (A) but at -3 a pyrimidine (C) occurs.

The DAS upstream region contains several candidates for the so called TATA-box (34). These sequences, indicated in Fig.6 by an open box, are located around the following positions: TTAAATATT at -399, AAAATATTA at -342 and TATAAATA at -52. The transcriptio-
Fig. 6  Nucleotide sequence of the DAS gene and its flanking sequence. 
In the 5'end region the 'TATA' boxes (open boxes), the region of multiple transcription start (conjugated arrows) and the TTT ... sequences (dashed line) are as indicated. Within the 3'end region the repeats of both the TA(A)AG sequence (dashed line above the sequence), and the AATAA sequence (solid line below the sequence) are indicated. In addition the sequence conforming to the consensus sequence T ... AAATAG ... A is indicated by a thicker solid line. The established 3'ends are indicated by the arrows.
Table 1

| Amino acid | A (%) | B (%) | Amino acid | A (%) | B (%) | Amino acid | A (%) | B (%) |
|------------|-------|-------|------------|-------|-------|------------|-------|-------|
| AsX        | 9.8   | 9.5   | GlX        | 10.3  | 11.5  | Val        | 6.4   | 6.5   |
| Thr        | 3.4   | 3.5   | Gly        | 8.0   | 8.3   | Ile        | 5.8   | 5.2   |
| Ser        | 5.5   | 5.5   | Ala        | 10.8  | 10.5  | Leu        | 8.0   | 7.9   |
| Tyr        | 5.2   | 4.4   | Lys        | 5.3   | 5.7   | Arg        | 4.9   | 3.9   |
| Phe        | 4.0   | 4.0   | His        | 2.5   | 2.4   | Pro        | 4.8   | 6.3   |

Note: Some of the amino acids in the protein (Met, Cys, Thr) were not analysed, Asn, Asp content as well as that of Glu, Gln were expressed as a sum; AsX and GlX respectively.

The amino acid composition derived from the presented reading-frame is similar to the data obtained from the analysis of the purified protein (kindly performed by Dr. L. Eggeling) as shown in Table 1. This confirms the correctness of the established reading frame and conclusively proves that the cloned gene encodes the DHAS protein. The determination of the N-terminal sequence of the protein was not possible, most probably because of modification.

DISCUSSION

In our study of the molecular basis of methanol utilization by
yeast, we have cloned and analysed the genes encoding the key enzymes involved in this complex and intriguing system. In the present paper, the DAS gene encoding the assimilatory dihydroxyacetone synthase is described. Our accompanying paper describes properties of a cloned MOX gene encoding the dissimilatory enzyme methanol oxidase (12).

We have sequenced the DAS coding region and its 5'- and 3'end flanking regions. The open reading frame of 702 amino acids, as presented in Fig.6, gives a protein of M_r of 77,000, a value in good agreement with the 78,000 daltons reported from SDS-PAGE analysis. The derived amino acid composition corresponds well to the data obtained by analysing the DHAS protein. This together with the results from hybrid selection analysis conclusively identifies the gene as DAS.

Interpretation of the hybrid selection data were complicated by the fact that mRNA isolated by hybrid selection gave two polypeptide bands on SDS-PAGE upon in vitro translation. Both bands were precipitable with specific anti-DHAS serum. Similarly, in vitro translation of total induced mRNA gave upon immunoprecipitation the same two bands, one of apparent M_r of genuine DHAS, the other obviously smaller than MOX. It seems that this situation is not a major artifact of the hybrid selection or/and the immunoprecipitation procedure since a careful examination on high resolution SDS-PAGE of translation products of total mRNA revealed the presence of three bands in the 70,000-80,000 daltons region; a genuine DHAS, a MOX and a third band. The latter band is not found in either crude protein extracts from induced cells nor in translation products of not induced mRNA. Our previous purification studies of DHAS and MOX protein (9) also did not show the existence of the third polypeptide in vivo. Thus we conclude that the observed band is an artifact of the in vitro translation system. This is supported by the observation that the strength of the third band, in all types of experiments (hybrid selection, immunoprecipitation, translation of total mRNA), strongly depends upon the in vitro translation conditions, such as amount of mRNA, reticulocytes batch and the addition of calf-liver tRNA.

Two 3'ends of transcription were mapped about 40 bp and 90 bp after the open reading frame. The 3'end flanking region bears se-
### Table 2 Codon usage in the DAS gene

| Codon  | ALA  | ARG  | ASN  | ASP  | CYS  | U  | G  |
|--------|------|------|------|------|------|----|----|
| U GA   | CCA  | CCA  | AAU  | GAU  | UGU  | 16| 9  |
| C     | 16*  | 29*  | 27*  | 9    | 9    |    |    |
| G     | 16** | 4    | 4    | 4    | 4    |    |    |
| AGA   | 21*  | 4    |      |      |      |    |    |
| GLN   | CAA  | GAA  | GGA  | CAU  | AUA  | 4 | 1 |
| G     | 23** | 36** | 12*  | 18*  | 15*  | 4 | 1 |
| GLU   | GAA  | GAA  | 9    | 9    | 9    | 9 | 9 |
| LEU   | UUA  | AAA  | MET  | AUG  | PHE  | 17 | 10|
| G     | 8    | 29*  | 31*  | 21*  | 21*  | 21*| 21*|
| LYS   | AAA  | 8    |     |     | UUU  | 17 | 7 |
| U     | 5    | 3    |    |    | 2    | 10| 3 |
| MET   | AUG  | 17   |    | 12*  | 22** | 22**|
| VAL   | GUA  | C  | 31*  | U  | 17   | 16|
| SER   | UCA  | THR  | TRP  | UUG  | TYR  | 6 | 6 |
| G     | 21** | 12** | 12** | 12** | 12** | 12**|
| LEU   | UUA  | AAA  |   |     | UUU  | 7 | 7 |
| C     | 5    | 5    | 7 |
| C     | 10   | 10   | 10 |
| C     | 33** | 33** | 33**|
| SER   | UCA  | THR  | 3   | UUG  | 6    | 6 |
| G     | 21** | 12** | 12**|
| AGU   |      |      | 5   | 2    | 2    | 2 |

* Strong bias in *S. cerevisiae.*
** Rarely used in *S. cerevisiae.*

Several pronounced features. Five repeats of TA(A)AG and four of (A)AATAA(A) sequences are located within the untranslatable region. A related AATAAA sequence is considered to be a polyadenylation signal in higher eukaryotes. Zaret and Sherman (35) however suggested that in *S. cerevisiae* this sequence alone does not function as such. Bennetzen and Hall (36) compared eleven *S. cerevisiae* 3'end regions and deduced a consensus sequence T...TAAATAAA...T...A...AT located some 30 bp from the mRNA 3'end. In fact in *H. polymorpha*, the TAAATAAG sequence flanked by Ts and As is found near the established 3'ends. The *H. polymorpha* 3'end region resembles the general architecture of the 3'end region of *S. cerevisiae* genes as proposed by Bennetzen and Hall (36) rather than the scheme suggested by Zaret and Sherman (35).

The codon usage in the DAS gene was compared with that of other organisms, in particular that of bacteria and yeast. Although DAS encodes an abundant protein, the codon bias is not so pronounced as has been found for *S. cerevisiae* i.e. codons are more randomly used. Out of 61 possible codons 56 are used in DAS (Table 2), whereas for instance the *S. cerevisiae* genes for alcohol dehydrogenase and glyceraldehyde-3-phosphate dehydrogenase
use only 33 and 29 codons, respectively. In the DAS gene some amino acids are encoded without preference (Ala and Ser), whereas others show a striking preference (Table 2). For amino acids like Arg (AGA) the bias is comparable with the S. cerevisiae situation. Such strong biases are denoted in Table 2 with a star (*). Mostly, however, the codons which are frequent in DAS are rarely or not at all used in S. cerevisiae. The typical examples are marked with a double star (**). Although we do not intend to draw too far reaching conclusions, we have the impression that the codon bias of DAS resembles more that of E. coli than that of Saccharomyces and is probably related to the higher GC content of H. polymorpha. As described in the accompanying paper, the MOX gene shows a very similar codon usage (12).

As we have reported previously (9) the cloned gene codes for an abundant protein. In this paper we have determined the DAS mRNA to constitute about 7% of the total polyA mRNA in cells grown on 1% methanol as the sole carbon source. This value is an underestimation since by the S1 mapping procedure only intact mRNA is measured and the fact that polyA mRNA preparations still contain some rRNA is not taken into account. Additional studies (Janowicz et al., in preparation) suggest that the high level of mRNA is due to a high transcription rate rather than to accumulation of the mRNA. We have ruled out the possibility of multiple gene copies (Janowicz et al., unpublished).

The DAS gene is strongly repressed by glucose. No trace of mRNA complementary to any region of the gene was found by Northern hybridization and S1 mapping, when cells were grown on 3% glucose. Studies on the functional role of the 5'end sequences of the DAS gene involved in this regulation are in progress.

REFERENCES
1. Tani, Y. (1984) Microbiology and Biochemistry of Methylotrophic Yeast in Methylotrophs: Microbiology, Biochemistry and Genetics, Hou, C.T., ed. pp.55-85, CRC Press, Boca Raton, USA
2. Sahm, H. (1977) Adv.Biochem.Eng. 6, 77-103.
3. Veenhuis, M., van Dijken, J.P. and Harder, W. (1983) Adv. Microbiol.Physiol. 24, 2-76.
4. Sahm, H. and Wagner, F. (1973) Eur.J.Biochem. 36, 250-256
5. Schütte, H., Flossdorf, J., Sahm, H. and Kula, M.R. (1976) Eur.J.Biochem. 62, 151-160.
6. Bystrykh, L.V., Sokolov, A.P. and Trotsenko, Y.A. (1981) FEBS Lett. 132, 324-328.
7. Kato, N., Higuchi, T., Sakazawa, C., Nishizawa, T., Tany, Y. and Yamada, H. (1982) Biochim. Biophys. Acta 715, 143-150.
8. Roa, M. and Blobel, G. (1983) Proc. Natl. Acad. Sci. USA 80, 6872-6876.
9. Roggenkamp, R., Janowicz, Z., Stanislawski, B. and Hollenberg, C.P. (1984) Mol. Gen. Genet. 194, 489-499.
10. van Dijken, S.P., Veenhuis, M. and Harder, W. (1982) Ann. N.Y. Acad. Sci. 386, 200-216.
11. Dobson, M.Y., Tuite, M.F., Roberts, N.A., Kingsman, A.S., Kingsman, S.M., Perkins, R.E., Lonroe, S.C., Dunbar, B. and Fothergill, L.A. (1982) Nucl. Acid Res. 10, 2625-2637.
12. Ledeboer, A.M., Maat, J., Visser, C., Bos, J.W., Verrips, C.T., Janowicz, Z., Eckart, M., Roggenkamp, R. and Hollenberg, C.P. (1985) Nucleic Acid Res., accompanying paper.
13. Struhl, K., Stinchcomb, D.T., Sherer, S. and Davis, R. (1979) Proc. Natl. Acad. Sci. USA 76, 1035-1039.
14. Birnboim, H.C. and Doly, J. (1979) Nucl. Acid Res. 7, 1513-1523.
15. Williams, B.G. and Blattner, F.R. (1979) J. Virol. 29, 555.
16. Fritsch, E.F. (1982) in Molecular Cloning, Cold Spring Harbor Manual, Maniatis, T., Fritsch, E.F. and Sambrook, J. eds. pp. 371-372.
17. Maniatis, T., Hardison, R.C., Lacy, E., Lauer, C., O'Connell, D., Quon, D., Sim, D.K. and Efstratiadis, A (1978) Cell 15, 687-701.
18. Hohn, B. (1979) Methods in Enzymology 68, 299-309.
19. Ceri, H. and Maeba, P.Y. (1973) Biochem. et Biophys. Acta 312, 337-348.
20. Goldbach, R.W., Borst, P., Bollen-de-Boer, J.E. and van Bruggen, E.F.J. (1978) Biochem. et Biophys. Acta 521, 169-186.
21. St. John, T.P. and Davis, R.W. (1979) Cell 16, 443-452.
22. Kramer, R.A. and Anderson, N. (1980) Proc. Natl. Acad. Sci. USA 77, 6541-6545.
23. Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
24. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
25. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161-178.
26. Barnes, W.M. and Bevan, M. (1983) Nucl. Acids Res. 11, 349-368.
27. Favarolo, J., Treisman, R. and Kamen, R. (1980) Methods in Enzymology 65, 718.
28. Bünemann, H., Westhoff, P. and Herrmann, R.G. (1982) Nucl. Acid Res. 10, 7163-7178.
29. Bünemann, H. (1982) Nucl. Acid Res. 10, 7179-7194.
30. Cleveland, D.W., Fischer, S.G., Kirchner, M.W. and Laemmli, U.K. (1977) J. Biol. Chem. 252, 1102-1106.
31. Thomas, P.S. (1980) Proc. Natl. Acad. Sci. USA 77, 5201-5205.
32. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
33. Kozak, M. (1981) Nucl. Acid Res. 9, 5233-5252.
34. Grannon, F., O'Hare, K., Perrin, F., Le Pennec, S.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garopin, A. and Chambon, P. (1979) Nature 278, 428-434.
35. Zaret, K.S. and Sherman, F. (1982) Cell 28, 563-573.
36. Bennetzen, J.H. and Hall, B.D. (1982) J. Biol. Chem. 257, 3018-3025.