

Identification and cloning of the *Neurospora crassa* glyceraldehyde-3-phosphate dehydrogenase gene, *gpd-1*

M. Sahni and J. A. Kinsey- Department of Microbiology, Molecular Genetics and Immunology, University of Kansas Medical Center, Kansas City, KS 66160

In work initially intended to use the *am* gene coding sequences as a reporter gene, 5' RACE PCR (Frohman *et al.*, 1988 Proc. Natl. Acad. Sci. USA. **85**:8998-9002) with three gene specific nested primers was performed. The product was cloned and sequenced, but found not to represent the *am* gene. Comparison to sequences in Genbank revealed that the product could encode a product homologous to glyceraldehyde-3-phosphate dehydrogenase (GPD) from a variety of other organisms. Consequently the PCR product was used to screen a lambda gt-11 expression library (Sachs *et al.* 1986 J. Biol. Chem **261**:869-873). The 1.3 kb insert from one cDNA clone was sequenced (Figure 1) and used to screen a *Neurospora* genomic library made in an EMBL-3 vector by E. Cambareri. All of the positive clones had a 7 kb *Bam*HI fragment. Relevant portions of one of the genomic clones was sequenced (Figure 1) revealing two introns. Although the complete genomic clone was not sequenced, comparison of restriction fragments from the cDNA and genomic clones indicated that no other introns are present in the *Neurospora gpd-1* gene.

Southern blot analysis of restriction enzyme digested DNA from Oak Ridge and Mauriceville strains revealed a polymorphism of *kpn*I sites at or near the *gpd-1* locus, allowing RFLP mapping using the small set of tester progeny as described by Metzenberg *et al.* (Metzenberg *et al.* 1984, *Neurospora* Newsl. **31**:35-39). The results shown in Table 1 indicate that *gpd-1* is located on linkage group IIR near the *arg-12* locus. Northern blot analysis using *gpd-1* cDNA as probe revealed a single strong band of 1.3 kb in length (data not shown).

One interesting question is how did we clone the *gpd-1* fragment by 5' RACE when we were using a nested set of three specific *am* primers for the amplification? When the sequence was analyzed it became apparent that each of the primers had 3' ends with five-to-six base pairs of perfect complementarity to sequences near the 5' end of the *gpd-1* message and that these sequences appeared in the same order in the *gpd-1* message as did the "specific" sequences in the *am* message. Given the abundance of *gpd-1* message this made amplification of the 5' end of the *gpd-1* gene probable during the 5' RACE experiment. Clones with either cDNA or genomic inserts are available from the Fungal Genetics Stock Center.

Table 1. RFLP mapping of *gpd-1a*.

GENE	11	12	13	14	15	16	17	18	19	20
<i>arg-12</i>	(O)	O	O	M	M	(M)	M	M	O	O
<i>gpd-1</i>	(O)	O	O	M	M	(M)	M	M	O	O
	21	22	23	24	25	26	27	28	29	30
<i>arg-12</i>	M	O	M	O	O	M	O	O	M	M
<i>gpd-1</i>	M	O	M	O	O	M	O	O	M	M

aA comparison of the segregation of the *gpd-1 Kpn*I RFLP with segregation data for *arg-12* which is located on LGIIR; strains numbered 11-30 represent FGSC strains 4411-4430. O or M in a particular strain indicates a fragment identical to that of the Oakridge or Mauriceville strain respectively. Strain 4411 is the Oak Ridge (O) parent and strain 4416 is the Mauriceville (M) parent.

CCCGGTGACG	GAGTGCTCTG	GCTGCTTGTT	GGGAATTGCC	GAGGCTCGCA	ACTGGAGCAG	60
TCAGCAATGT	CAGCATCGAC	ATGTTCAAGT	TGACTCATTT	CAGTTGGTAT	TACAAAGACT	120
GAACCCGTGA	AGCACATAGC	GTGACCGAAT	CACGGATTCT	CCGGCAAGGA	GCTTGTTTCA	180
TTGTTGCCTC	TTGTCGGCGG	CTTTCAAAGC	AAAAAAGGAT	GGGAATCTCT	TCATGCCAAG	240
GGCGCGGCCG	AGTACTGCGC	TAACACTAGA	CGCCAAGCCA	TTGGAGAGTG	GCCCCACCTC	300
ATCCCACCAT	GTCCCACCAC	CACAGCCCAC	CATGGAGCAA	AGCGTATGAT	GCAACCACGA	360
TGGGAGGCGG	CTGGTGGGAT	GGAAGGAACG	AGCAAAACCA	CCCACCCATT	GACCACCCCA	420

CCCTCAAACC AAATTTATGT CGCTCATGCC ACCACGGTGA CATTTGGCAG GCATTGAGAG 480
 CGTTCAGGGG GGTGATGAGG AGCTCCCCTC CTCTTTTGCC CCTCCTTGCC GACTGGGGAT 540
 TACCACAGGC TGATAACCAG ACTGGACGCG AGCAGGGCAG CTGGAGTCGG CTGGGAAACT 600
 AGATAATAGA TAGTACAAGA ATCTCCTCCT GCCTCCCAAC TTTTCTCTTT CTTTCTCTTG 660
 CTTTCATCATC ATCCTCGCGA TACCAAGTTC ACTTCCAACC AAAACCTTC TTCCAAACCA
 720
 -----intron I-----
 CATCAGGTAT GTTGTGACTG CCCTCGCATT TACAGAAACC GAGCTTCCTT CCTCAACACT 780

 TCCAATCATC GTCACCTCCC TTGTCAGCGG CGGCGGCAGC AGCAGCAGTA GCAGAAGCAG 840

 AAGCAGAAGC AGCAGCTACC CCGCACCTTC CTGACCCCGT CCCGACCCCG TCCCATCTCA 900

 TCCTCAGTCA GTTCCTCCCG CCTCGCTGCC AAGCTGCGCA CAGCATCTGG TGTCTGCGTC 960

 TGTTTCCCCC CAAGAGGAAG TGGACGAGAC TCAGATCGGA CTGGCATGGA TGCTGGTGGT 1020

 GGTGGCGGCA TTGGAAGGGT TCCTCGGAAT CGCTCCTCCC CGATCCTACC TGCAGTCGGT 1080

 CCCTCCGTGT TTTGGGCGCT CCTCGTGTCC AATTGTTCTG CCACGCAAAC ATGTGAACAG 1140

 ACGAGACCGA ACAGGATAAG GAAGGGCAGG CAGACGAGTC CGGCTTTAAA ACCCAGACTT 1200

 TCCTTCATCC TACCACTCAT CATCATCTTA CAACCTTCAA CAACTTGCTT CACAAGGTCT 1260

 TGATACTTAC TCGTCTTCAC TCCAACAGTC AAC ATG GTC GTC AAG GTC GGC ATC
 AAC 1318
 M V V K V G I N
 GGT TTC GGC CGT ATC GGT CGC ATT GTC TTC CGC AAT GCC ATT GAG CAC GAT GAC
 1371
 G F G R I G R I V F R N A
 I E H D D

 ATC CAC ATC GTC GCT GTC AAC GAC CCC TTC ATT GAG CCC AAG TAC GCT GTAAGTT
 1425
 I H I V A V N D P F I E P
 K Y A
 -----intron 2-----
 GGCC TCGCTCACAT AGATCCCTTG TCTCATATGACAACTCAGAC TCTGACCATC ATCCCT 1486

 CTTA CAG GCT TAC ATG CTC CGC TAC GAC ACC ACC CAC GGC AAC TTC AAG GGC ACC
 1541
 A Y M L R Y D T T H G
 N F K G T
 ATC GAG GTT GAC GGT GCT GAC CTC GTC GTC AAC GGC AAG AAG GTC AAG TTC TAC 1595
 I E V D G A D L V V N G K
 K V K F Y
 ACT GAT GCC GAC CCC GCT GCC ATC CCC TGG TCC GAG ACC GGT GCC GAC TAC ATT 1649
 T D A D P A A I P W S E T
 G A D Y I
 GTC GAG TCC ACT GGT GTC TTC ACC ACC ACC GAG AAG GCC TCC GCC CAC TTG AAG
 1703
 V E S T G V F T T T E K A
 S A H L K
 GGT GGT GCC AAG AAG GTC ATC ATC TCT GCC CCC TCT GCT GAT GCC CCC ATG TAC 1757
 G G A K K V I I S A P S A
 D A P M Y
 GTT ATG GGT GTC AAC AAC GAG ACC TAC GAT GGC TCC GCC GAC GTC ATC TCC AAC 1811
 V M G V N N E T Y D G S A D
 V I S N
 GCC TCT TGC ACC ACC AAC TGC TTG GCT CCC CTC GCC AAG GTC ATC CAC GAC AAC 1865
 A S C T T N C L A P L A K
 V I H D N

TTC	ACC	ATC	GTC	GAG	GGT	CTC	ATG	ACC	ACC	GTC	CAC	TCC	TAC	ACC	GCC	ACC	CAG	1919
F		T	I	V	E		G	L		M	T	T	V	H	S			
Y	T	A	T	Q														
AAG	ACC	GTC	GAT	GGT	CCT	TCC	GCC	AAG	GAC	TGG	CGC	GGT	GGC	CGC	ACT	GCT	GCT	1973
K	T	V	D	G	P	S	A	K	D	W	R	G						G
R	T	A	A															
CAG	AAC	ATC	ATT	CCC	AGC	AGC	ACT	GGT	GCC	GCC	AAG	GCC	GTC	GGC	AAG	GTC	ATC	2027
Q	N	I	I	P	S	S	T	G	A	A	K	A						
V	G	K	V	I														
CCC	GAC	CTC	AAC	GGC	AAG	CTC	ACT	GGT	ATG	GCC	ATG	CGT	GTC	CCC	ACC	GCC	AAC	2081
P	D	L	N	G	K	L	T	G	M	A	M	R	V					
P	T	A	N															
GTC	TCC	GTT	GTC	GAT	CTT	ACT	GCC	CGC	ATC	GAG	AAG	GGT	GCT	ACC	TAC	GAT	GAG	2135
V	S	V	V	D	L	T	A	R	I	E	K	G						
A	T	Y	D	E														
ATC	AAG	GAG	GTC	ATC	AAG	AAG	GCC	TCT	GAG	GGT	CCC	CTC	GCT	GGC	ATC	CTT	GCC	2189
I	K	E	V	I	K	K	A	S	E	G	P	L						
A	G	I	L	A														
TAC	ACC	GAG	GAT	GAG	GTT	GTC	TCT	TCC	GAC	ATG	AAC	GGC	AAC	CCC	GCC	TCC	TCC	2243
Y	T	E	D	E	V	V	S	S	D	M	N	G	N					
P	A	S	S															
ATC	TTC	GAT	GCC	AAG	GCT	GGT	ATC	TCC	CTC	AAC	AAG	AAC	TTC	GTC	AAG	CTT	GTC	2297
I	F	D	A	K	A	G	I	S	L	N	K	N	F					
V	K	L	V															
TCC	TGG	TAC	GAC	AAC	GAG	TGG	GGC	TAC	TCT	CGC	CGT	GTC	CTC	GAC	CTC	ATC	TCC	2351
S	W	Y	D	N	E	W	G	Y	S	R	R	V	L					
D	L	I	S															
TAC	ATC	TCC	AAG	GTC	GAT	GCC	AAG	AAG	GCT	TAA	ATCGGT	TGCGTACCCGCACGGTTA						2408
Y	I	S	K	V	D	A	K	K	A									
TG	AAGTAATGGT	CTTTTCCTAG	ATATGAAGAA	AAAAAAAGGG	CAATGATTCC	GTGGGATT												2468
GA	ACTCGAGCAT	GTTGGATCTC	GGGCAGTCCT	GCTTAAAGTA	AAATAATATC	CGAACTCAA												2528
ATAG	ATACCAAGTTCACTTCG																	2552

Figure 1. Sequence of the *gpd-1* gene. The sequence presented represents a combination of sequences from cDNA and genomic DNA. The first nucleotide of the cDNA sequenced is at 677. This is 5 nucleotides downstream of a consensus fungal transcriptional start site at position 666-673 (Bruchez *et al.* 1993 Fungal Genet. Newsl. **40**:89-96). The pyrimidine box characteristically found upstream of the transcriptional start sites of fungal genes is underlined. The two introns are indicated by dashed overlining. There was no polyadenylated tract in the cDNA sequenced