Volume 241, number 1,2, 83–88 FEB 06552 December 1988

The *Saccharomyces cerevisiae ARO1* gene

An example of the co-ordinate regulation of five enzymes on a single biosynthetic pathway

Kenneth Duncan*, R. Mark Edwards+ and John R. Coggins

Department of Biochemistry, University of Glasgow, Glasgow G12 8QQ, Scotland and +*Searle Research and Development, Lane End Road, High Wycombe, Bucks HP12 4HL, England*

Received 20 September 1988

The *ARO1* gene of *Saccharomyces cerevisiae* encodes the arom multifunctional enzyme. Specific inhibitors of amino acid biosynthesis have been used to obtain evidence that expression of a cloned *ARO1* gene is regulated in response to amino acid limitation. Northern blot analysis and sequence studies indicate that *ARO1* is regulated by the well characterised *S. cerevisiae* 'general control' mechanism. This provides a very economical means of simultaneously tailoring the synthesis of five shikimate pathway enzymes to the needs of the cell.

Shikimate pathway; Arom multifunctional enzyme; ARO1 gene regulation; (*Saccharomyces cerevisiae*)

1. INTRODUCTION

The product of the *Saccharomyces cerevisiae ARO1* gene, the arom protein, is a multifunctional enzyme which catalyses five consecutive steps on the shikimate, or common aromatic pathway. This pathway is responsible for the biosynthesis of chorismate from the carbohydrate precursors erythrose-4-phosphate and phospho*enol*pyruvate. One remarkable feature of the pathway is the diversity of organisation of the enzymes and the genes encoding them in a wide variety of species. The pentafunctional arom enzyme represents one extreme-one gene, five activities [1,2]; at the opposite extreme, *Escherichia coli* has five widely scattered genes [3], from which five separable products are synthesised [4,5]. Arom genes have been cloned from a number of organisms, including *S. cerevisiae* [6], *Aspergillus nidulans* [7], *Schizosaccharomyces pombe* [8] and *Neurospora crassa* [9].

We report here that in *S. cerevisiae*, the five shikimate pathway enzymes encoded by the *ARO1* gene are co-ordinately regulated in response to amino acid limitation by the so-called 'general control' mechanism [10–13].

Correspondence address: J.R. Coggins, Dept of Biochemistry, University of Glasgow, Glasgow G12 8QQ, Scotland

* *Present address:* Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 25 Shattuck Street, Boston, MA 02115, USA

+ *Present address:* British Biotechnology, Brook House, Watlington Road, Oxford OX4 5LY, England

2. MATERIALS AND METHODS

2.1. *Materials*

[α-^{32}P]dCTP was purchased from Amersham International, Amersham, Bucks. RNase-free DNase I was from BCL, Lewes, East Sussex. Histidine, aminotriazole and 5-methyltryptophan were purchased from Sigma, Poole, Dorset. Analytical grade glyphosate was a gift from Dr Stuart Ridley, ICI, Jealot's Hill, Berks. Schleicher & Schull nitrocellulose (0.22 μM) was obtained from Anderman & Co., Kingston-on-Thames, Surrey.

2.2. *DNA cloning and sequencing*

Recombinant DNA manipulations were carried out as described in [14]; DNA sequencing by the M13/dideoxy method was as described in [15].

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

2.3. *Isolation of RNA*

1 ml samples of an exponentially growing *S. cerevisiae* culture were harvested by centrifugation (microfuge, 1 min) and resuspended in 200 μl of 200 mM Tris-HCl (pH 7.5), 0.5 M NaCl, 10 mM EDTA. 200 μl of phenol/chloroform (1:1) was added, along with glass beads (0.5 g), and the sample disrupted by vortexing (2 × 30 s bursts). The aqueous layer was removed, and nucleic acids were precipitated following addition of ethanol. After recovery, nucleic acids (mostly RNA) were resuspended in 9 μl of 10 mM Tris-HCl (pH 8.0), 1 mM EDTA; 1 μl of RNase-free DNase I was added and the sample incubated (15 min at 37°C).

2.4. *Northern blot analysis of RNA*

DNase-treated, formamide-denatured RNA was separated by electrophoresis through a 1% agarose gel containing formaldehyde, blotted to nitrocellulose, and probed as described in [14]. Probe DNA was prepared by second-strand synthesis of an M13mp19 recombinant template which carries a 2.6 kb *Bam*HI-*Sst*I fragment from pFL6. In the presence of [α-^{32}P]dCTP, the strand complementary to the *ARO1* message was labelled.

2.5. *Preparation of cell extracts for enzyme assays*

50 ml *S. cerevisiae* S288C cultures (grown on 0.67% yeast minimal medium without amino acids, containing 2% glucose) were supplemented with (final concentration): histidine (0.3 mM) throughout growth; aminotriazole (10 mM), 5-methyltryptophan (0.5 mM) and glyphosate (10 mM), when A_{600} = 0.5. After a further 3 h growth, cells were harvested by centrifugation and washed (8 ml distilled water). The cell paste was resuspended in 350 μl of 100 mM Tris-HCl (pH 7.5), 1.4 mM β-mercaptoethanol, 1.2 mM phenylmethanesulphonyl fluoride; glass beads were added until the liquid was covered and the cells were disrupted by vortexing (3 × 30 s bursts). 1.5 ml of the same buffer was added and the suspension vortexed for a further 30 s. The supernatant was removed and centrifuged (200000 × *g* for 2 h), to produce the cleared crude extract for assay. Extracts were prepared from six identical flasks for each supplement.

2.6. *Enzyme assays*

3-Dehydroquinase (3-dehydroquinate hydrolyase, EC 4.2.1.10) was assayed in the forward direction as described in [16]. Shikimate dehydrogenase (Shikimate:NADP$^+$ oxidoreductase, EC 1.1.1.25) was assayed in the reverse direction as described in [16]. Glutamate dehydrogenase (L-glutamate:NADP$^+$ oxidoreductase [transaminating], EC 1.4.1.13) was assayed as described in [17], monitoring the conversion of α-ketoglutarate to glutamate. Specific activity is in $\mu mol^{-1} \cdot min^{-1} \cdot mg^{-1}$.

3. RESULTS

3.1. *DNA sequence analysis of* ARO1

DNA sequence analysis was carried out on two independently isolated *ARO1* clones, pFL6 and pME173. pFL6 is a derivative of YpAr1 [6]; pME173 carries 6.5 kb of *S. cerevisiae* genomic DNA, and was isolated from a *Sau*3A partial digest library (Dacey, S.A. and Edwards, R.M., unpublished). The sequencing strategy used has been outlined in [15], and consisted of a combination of sequencing restriction enzyme generated DNA fragments cloned into M13 and specific priming using synthetic oligonucleotides. In total, 6489 bp of the DNA sequence were obtained and part of this is shown in fig.1. Translation of this DNA sequence revealed a single, long open reading frame of 1588 amino acids from the first ATG/methionine to the stop codon. Computer analysis of the open reading frame revealed extensive homologies at the amino acid level with the sequence of the arom protein of *A. nidulans* [18,19], and with the sequences of the corresponding individual *E. coli* enzymes [15].

Inspection of the 5′-flanking sequence of *ARO1* reveals a number of features typically found upstream of *S. cerevisiae* genes. Immediately before the proposed *ARO1* translation start is an A at the third position before the ATG codon. An A at this position is invariant amongst *S. cerevisiae* genes [20]. The location of the transcription start site (5′-end of the message) in the DNA sequence was determined by the detection of primer extensions with reverse transcriptase on *S. cerevisiae* mRNA, and confirmed by detection of DNA/RNA hybrids protected from digestion by nuclease S_1 (not shown). The 5′-terminus is the A at position −93 (fig.1, located by *). Near to this (positions −92 to −86) is the sequence CCATTAG which is similar to the consensus PyCATTCPu eukaryotic gene expression signal for a 'capping site' [21]. A number of regions in the 5′-upstream sequence show homology to the 'TATA', or 'Golberg-Hogness' box which is thought to be an essential RNA polymerase II recognition sequence within the eukaryotic promoter. Sequences similar to the consensus TATAAAA can be found at positions −138 to −132 and −119 to −113. TATA boxes are usually found in higher eukaryotes at about 35 bp before the transcription start site, but in *S. cerevisiae* genes the distance is somewhat variable and usually longer [22]. Another common feature is the 'CAAT box' with the consensus GCC/TCAATCT, which has been found in several eukaryotic genes about 80 bp upstream of the RNA transcription start site [23]. An almost perfect match to this consensus is found between

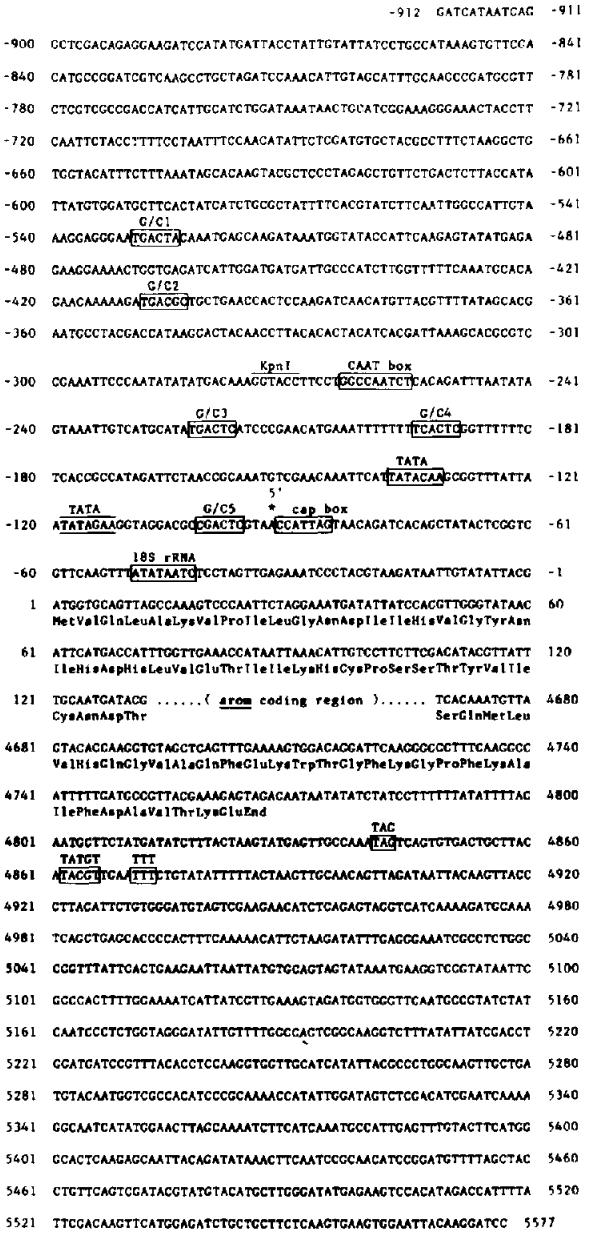

Fig.1. DNA sequences flanking the *AROI* gene.

positions −264 and −256, making a spacing of 163 bp before the transcription start site. At positions −50 to −43 the sequence ATATAATC is found. This shows partial complementarity with the 3′-terminus of *S. cerevisiae* ribosomal RNA (3′-AUUACUAG-5′) [24].

By Northern blot analysis (fig.2, and unpublished data), we have estimated that the length of the polyadenylated *AROI* mRNA is 5050 nucleotides. Thus the true length of the messenger is around 5000 nucleotides after the removal of the poly(A) tail, which is typically 50 nucleotides long in *S. cerevisiae* [25]. On the basis that the mRNA 5′-end is at position −93, the 3′-terminus is therefore located at approximately position 4907. Zaret and Sherman [26] have proposed the tripartite transcription termination and polyadenylation signal sequence TAG...TAGT or TATGT..(AT rich)..TTT. Poly(A) addition occurs 10 to 40 bp downstream from the TAGT part of this signal. One such sequence is found in the *AROI* 3′-flanking DNA, positions 4841 to 4873 (fig.1), and this is located at the appropriate distance upstream of the deduced mRNA terminus.

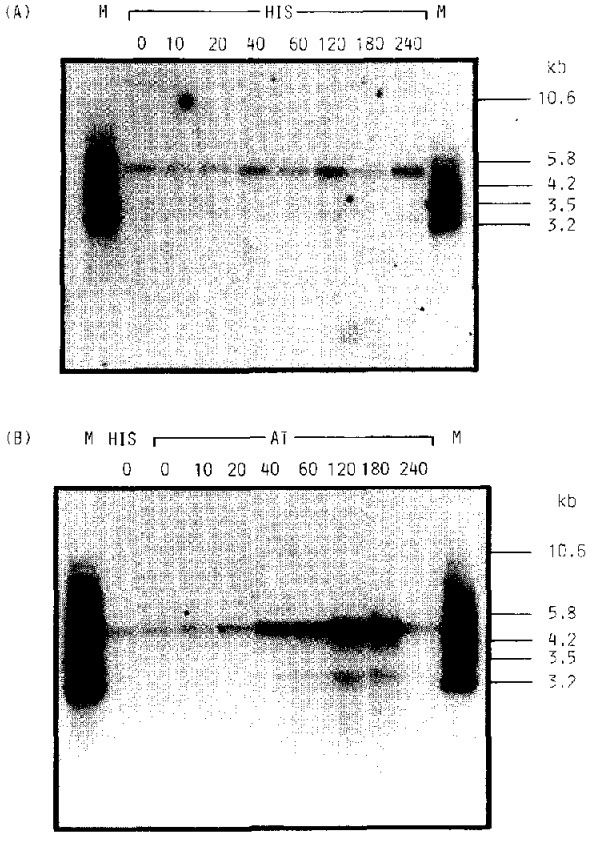

Fig.2. Northern blot of mRNA isolated from *S. cerevisiae*, and probed for *AROI* sequences. Numbers indicate the time in minutes. In A, the culture was supplemented with histidine; in B, aminotriazole was added at time point 0. M indicates marker DNA fragments of pFL6.

3.2. *General control and* ARO1 *expression*

The most striking feature in the 5′-flanking sequence is the presence of multiple copies of the hexanucleotide TGACTC. It has recently been shown that a number of co-regulated genes under general control have this characteristic short repeated DNA sequence in the 5′-noncoding region [27–31]. TGACTC is the most highly conserved part of a longer consensus for the repeated sequence (AA/TGTGACTC) [27], and is important in vivo for binding of a trans-acting regulatory factor, the product of the *GCN4* gene [32]. Upstream of *ARO1*, five sequences which are homologous to the short consensus are present; those with a 6/6 or a 5/6 match are shown in fig.1. At least two copies of the sequence must be present for the gene to be under general control [27]. The repeat is normally followed by a run of Ts (which are necessary for binding of the GCN4 protein [32]). The repeat G/C3 (fig.1), which is a perfect match to the consensus sequence, is separated by 15 bp from the run of Ts at positions −184 to −191. There then follows an imperfect repeat, G/C4, which is itself followed 2 bp later by a run of 6 Ts. On the basis that these sequence homologies were located upstream of the putative *ARO1* TATA box(es), it was predicted that *ARO1* is under general control.

In order to test this hypothesis, the specific activities of 3-dehydroquinase, shikimate dehydrogenase, and the NADPH-dependent glutamate dehydrogenase were monitored under various growth conditions. *S. cerevisiae* was grown to exponential phase and a variety of supplements added. Growth was continued for a further 3 h before the cells were harvested and enzyme activities determined (table 1). The results show that the presence of aminotriazole (an inhibitor of the *S. cerevisiae HIS3* gene product, imidazoleglycerol phosphate dehydratase, step 7 of the histidine biosynthetic pathway) in the medium leads to a 2.5–3-fold increase in the specific activity of both 3-dehydroquinase and shikimate dehydrogenase, compared with cells grown in the presence of histidine. A similar result was obtained with cells grown in the presence of 5-methyltryptophan, which acts as a feedback inhibitor of anthranilate synthase, the first enzyme on the tryptophan biosynthetic pathway. No such increase was found in the levels of glutamate dehydrogenase under these conditions: the *GDH1* gene which encodes this enzyme is not regulated by the general control mechanism [11]. Both aminotriazole and 5-methyltryptophan are known to bring about induction of all of the enzymes under general control. A specific shikimate pathway inhibitor was also tested under the same conditions. The herbicide glyphosate (*N*-phosphonomethylglycine) acts in plants by inhibiting another of the arom activities, 5-enolpyruvylshikimate-3-phosphate synthase [33]: both the bacterial and fungal enzymes are similarly affected [34,35]. In this case, a smaller increase in the specific activities of 3-dehydroquinase and shikimate dehydrogenase was observed in response to this inhibitor. It is not known whether this increase is a general control response, or an *ARO1*-specific response.

Table 1

Growth supplement	3-Dehydro-quinase		Shikimate dehydrogenase		Glutamate dehydrogenase	
	SA	RA	SA	RA	SA	RA
Histidine	3.54	1	0.022	1	0.492	1
Aminotriazole	9.08	2.56	0.072	3.27	0.694	1.41
5-Methyl-tryptophan	9.39	2.65	0.067	3.05	0.658	1.33
Glyphosate	6.05	1.71	0.041	1.86	0.559	1.14

SA, specific activity; RA, relative activity

We have shown that the specific activity increase was due to de novo enzyme synthesis by following the specific induction of *ARO1* mRNA after starvation for amino acids. *ARO1* mRNA was detected on Northern blots during a time course of the induction, as detailed in section 2. Although no effort was made to quantify the total RNA loaded on the gel, all the samples were treated in precisely the same manner. It is clear that in the presence of aminotriazole in the growth medium, the level of the *ARO1* message increases dramatically, reaching a maximum after 3 h (fig.2). In cells grown in the presence of histidine, the *ARO1* mRNA level remains constant.

4. DISCUSSION

The product of the *S. cerevisiae ARO1* gene is a very large (M_r 174555) and complex multidomain, multifunctional enzyme [15]. Although a number

of multifunctional enzymes have been identified in *S. cerevisiae*, this protein is unique in its ability to catalyse five consecutive steps out of seven on a single biosynthetic pathway. Several reasons have been postulated for the existence of this form of enzyme organisation, typical of the fungi [36]. These include catalytic facilitation [37], channeling [38], and co-ordinate activation or regulation of enzyme activity [39]. There is at present no conclusive evidence that any of these features can be ascribed to arom (Nimmo, G.A., Boocock, M.R., Lambert, J.M. and Coggins, J.R., unpublished), leaving unanswered the question of what advantage is to be gained by multifunctional gene organisation. One consequence of the multifunctional gene organisation is that the enzyme activities are synthesised in stoichiometric amounts. This avoids the problem of co-ordinating the expression of five separate enzyme genes. The next logical step would then be to tailor synthesis of the multifunctional protein to the biochemical needs of the cell and thus avoid the wasteful synthesis of unneeded enzyme. *S. cerevisiae* achieves this by placing *ARO1* under the general control mechanism, making it responsive to aromatic amino acid limitation. In connection with this co-ordination of gene expression it is interesting to note that in *N. crassa*, where the arom protein has been kinetically characterised [40], the turnover numbers of the five enzymes are very similar.

Acknowledgements: We thank Frank Larimer (Oak Ridge) for the generous gift of pFL6. Work carried out in Glasgow was supported by research grants from Searle Research and Development and the Science and Engineering Research Council.

REFERENCES

[1] Giles, N.H., Case, M.E., Partridge, C.W.H. and Ahmed, S.I. (1967) Proc. Natl. Acad. Sci. USA 58, 1453–1460.
[2] De Leeuw, A. (1968) Genetics 56, 554–555.
[3] Bachmann, B. (1983) Microbiol. Rev. 44, 180–230.
[4] Berlyn, M.B. and Giles, N.H. (1969) J. Bacteriol. 99, 222–230.
[5] Coggins, J.R., Boocock, M.R., Campbell, M.S., Chaudhuri S., Lambert, J.M., Lewendon, A., Mousdale, D.M. and Smith, D.D.S. (1985) Biochem. Soc. Trans. 13, 299–303.
[6] Larimer, F.W., Morse, C.C., Beck, A.K., Cole, K.W. and Gaertner, F.H. (1983) Mol. Cell. Biol. 3, 1609–1614.
[7] Kinghorn, J.R. and Hawkins, A.R. (1982) Mol. Gen. Genet. 186, 145–152.
[8] Nakanishi, N. and Yamamoto, M. (1984) Mol. Gen. Genet. 195, 164–169.
[9] Catcheside, D.E.A., Storer, P.J. and Klein, B. (1985) Mol. Gen. Genet. 199, 446–451.
[10] Schurch, A., Miozzari, J. and Hutter, R. (1974) J. Bacteriol. 117, 1131–1140.
[11] Wolfner, M., Yep, D., Messenguy, F. and Fink, G.R. (1975) J. Mol. Biol. 96, 273–290.
[12] Jones, E.W. and Fink, G.R. (1982) in: The Molecular Biology of the Yeast *Saccharomyces*. Metabolism and Gene Expression (Strathern, J.N. et al. eds) pp.181–299, Cold Spring Harbor, NY.
[13] Fink, G.R. (1986) Cell 45, 155–156.
[14] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, NY.
[15] Duncan, K., Edwards, R.M. and Coggins, J.R. (1987) Biochem. J. 246, 375–386.
[16] Coggins, J.R., Boocock, M.R., Chaudhuri, S., Lambert, J.M., Lumsden, J., Nimmo, G.A. and Smith, D.D.S. (1987) Methods Enzymol. 142, 325–341.
[17] Doherty, D. (1970) Methods Enzymol. 17A, 850–856.
[18] Charles, I.J., Keyte, J.W., Brammar, W.J. and Hawkins, A.R. (1985) Nucleic Acids Res. 13, 8119–8128.
[19] Charles, I.J., Keyte, J.W., Brammar, W.J., Smith, M. and Hawkins, A.R. (1986) Nucleic Acids Res. 14, 2201–2213.
[20] Dobson, M.J., Tuite, M.F., Roberts, N.A., Kingsman, A.J., Kingsman, S.M., Perkins, R.E., Conroy, S.C., Dunbar, B. and Fothergill, L.A. (1982) Nucleic Acids Res. 10, 1625–2637.
[21] Sures, I., Lowry, J. and Kedes, L.H. (1978) Cell 15, 1033–1044.
[22] Sentenac, A. and Hall, B. (1982) in: The Molecular Biology of the Yeast *Saccharomyces*. Metabolism and Gene Expression (Strathern, J.N. et al. eds) pp.561–606, Cold Spring Harbor, NY.
[23] Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1982) Nucleic Acids Res. 8, 127–142.
[24] Shine, J. and Dalgarno, L. (1974) Biochem. J. 141, 609–615.
[25] McLaughlin, C.S., Warner, J.R., Edmonds, M., Nakazato, H. and Vaughan, M.H. (1973) J. Biol. Chem. 248, 1466–1471.
[26] Zaret, K.S. and Sherman, F. (1982) Cell 28, 563–573.
[27] Hinnebusch, A.G. and Fink, G.R. (1983) J. Biol. Chem. 258, 5238–5247.
[28] Struhl, K. (1982) Nature 300, 284–287.
[29] Zalkin, H. and Yanofsky, C. (1982) J. Biol. Chem. 257, 1491–1500.
[30] Aebi, M., Furter, R., Prantl, F., Niederberger, P. and Hutter, R. (1984) Curr. Genet. 8, 165–172.
[31] Zalkin, H., Puluh, J.L., Van Cleemput, M., Moye, W.S. and Yanofsky, C. (1984) J. Biol. Chem. 259, 3985–3992.
[32] Hope, I.A. and Struhl, K. (1985) Cell 43, 177–186.
[33] Amhrein, N., Schab, J. and Steinrucken, H.C. (1980) Naturwiss. 67, 356–357.
[34] Boocock, M.R. and Coggins, J.R. (1983) FEBS Lett. 154, 127–133.

[35] Lewendon, A. and Coggins, J.R. (1983) Biochem. J. 213, 187–191.
[36] Coggins, J.R. and Boocock, M.R. (1986) in: Multifunctional Proteins – Structure and Evolution (Hardie, D.G. and Coggins, J.R. eds) pp.259–281, Elsevier, Amsterdam.
[37] Gaertner, F.H., Ericson, M.C. and DeMoss, S.A. (1970) J. Biol. Chem. 245, 595–600.
[38] Welch, G.R. and Gaertner, F.H. (1975) Proc. Natl. Acad. Sci. USA 72, 4218–4222.
[39] Welch, G.R. and Gaertner, F.H. (1976) Arch. Biochem. Biophys. 172, 476–489.
[40] Lambert, J.M., Boocock, M.R. and Coggins, J.R. (1985) Biochem. J. 226, 817–829.