

Characterization of the cDNA and Gene Coding for the Biotin Synthase of *Arabidopsis thaliana*¹

Lisa M. Weaver, Fei Yu, Eve Syrkin Wurtele, and Basil J. Nikolau*

Department of Biochemistry and Biophysics (L.M.W., B.J.N.) and Department of Botany (F.Y., E.S.W.), Iowa State University, Ames, Iowa 50011

Biotin, an essential cofactor, is synthesized de novo only by plants and some microbes. An *Arabidopsis thaliana* expressed sequence tag that shows sequence similarity to the carboxyl end of biotin synthase from *Escherichia coli* was used to isolate a near-full-length cDNA. This cDNA was shown to code for the *Arabidopsis* biotin synthase by its ability to complement a *bioB* mutant of *E. coli*. Site-specific mutagenesis indicates that residue threonine-173, which is highly conserved in biotin synthases, is important for catalytic competence of the enzyme. The primary sequence of the *Arabidopsis* biotin synthase is most similar to biotin synthases from *E. coli*, *Serratia marcescens*, and *Saccharomyces cerevisiae* (about 50% sequence identity) and more distantly related to the *Bacillus sphaericus* enzyme (33% sequence identity). The primary sequence of the amino terminus of the *Arabidopsis* biotin synthase may represent an organelle-targeting transit peptide. The single *Arabidopsis* gene coding for biotin synthase, *BIO2*, was isolated and sequenced. The biotin synthase coding sequence is interrupted by five introns. The gene sequence upstream of the translation start site has several unusual features, including imperfect palindromes and polypyrimidine sequences, which may function in the transcriptional regulation of the *BIO2* gene.

Biotin is an essential cofactor for a variety of carboxylases and decarboxylases found in diverse metabolic pathways of all organisms (Knowles, 1989). Despite the ubiquitous requirement for biotin, its de novo synthesis is restricted to plants and some microbes. Biotin biosynthesis has been best studied in bacteria. In *Escherichia coli*, biotin synthesis requires at least six genes and proceeds via a four-step pathway from the precursors pimelic acid and Ala (Eisenberg, 1987). Recent data indicate that biotin may be synthesized by a very similar route in plants (Schneider et al., 1989; Shellhammer and Meinke, 1990; Baldet et al., 1993b).

Biotin is composed of an imidazolidone ring fused to a valeric acid-substituted thiophene ring. The formation of the thiophene ring is the final process in biotin biosynthesis and involves the insertion of a sulfur atom into dethiobiotin to form biotin (Fig. 1). The direct donor of the sulfur atom used in this reaction may be Cys (Birch et al., 1995). In

E. coli, this final reaction in biotin biosynthesis requires the product of the *bioB* gene, biotin synthase (Eisenberg, 1987). Recent reports indicate that additional proteins, other than the *bioB* gene product, may also be required in the catalysis of this reaction (Ifuku et al., 1992; Sanyal et al., 1994; Birch et al., 1995). Furthermore, it is not clear if this reaction occurs via the formation of the intermediate 9-mercaptodethiobiotin or in one step (Baldet et al., 1993b).

Until recently, very little was known about biotin biosynthesis in plants. This is particularly remarkable considering the fact that plants are a major source of biotin in the biosphere. Precursor feeding studies with *bio1*, a biotin auxotrophic mutant of *Arabidopsis thaliana* (Schneider et al., 1989; Shellhammer and Meinke, 1990), and later biochemical fractionation of plant extracts (Baldet et al., 1993b) indicate that in plants biotin is probably synthesized by a pathway similar to that found in bacteria. In this paper, we report the structure of the *Arabidopsis* cDNA and gene (*BIO2*) coding for biotin synthase. To our knowledge, this is the first plant biotin biosynthetic gene to be characterized. The data presented below indicate that in addition to the homologous nature of the biotin biosynthetic pathways in plants and bacteria, the structure of the enzymes involved may also be homologous.

MATERIALS AND METHODS

Materials

Seeds of *Arabidopsis thaliana* (L.) Heynh., ecotype Columbia, were germinated in sterile soil, and plants were grown in a growth room maintained at 25°C with a 15-h day period. The EST 86e12 (Newman et al., 1994) was obtained from the *Arabidopsis* Biological Resource Center (Ohio State University, Columbus). A cDNA library in the vector λ gt10, prepared from poly(A)⁺ RNA isolated from developing siliques of *Arabidopsis*, was a gift from Dr. David W. Meinke (Oklahoma State University, Stillwater). An *Arabidopsis* genomic library made in the vector λ FIX, from DNA isolated from the ecotype *Landsberg erecta* (Voytas et al., 1990), was obtained from *Arabidopsis* Biological Resource Center. The *Escherichia coli* strain R875 described by Cleary and Campbell (1972) was obtained from the *E. coli* Strain Center (Yale University, New Haven, CT). The genotype of this strain is *bioB17, lac-3350, LAM-, IN(rrnD-rrnE)1*.

¹ Supported in part by funds provided by the state of Iowa. This is journal paper No. J-16511 of the Iowa Agriculture and Home Economics Experiment Station, Ames, IA, project Nos. 2913 and 2997.

* Corresponding author; e-mail dimmas@iastate.edu; fax 1-515-294-0453.

Abbreviations: EST, expressed sequence tag; ORF, open reading frame.

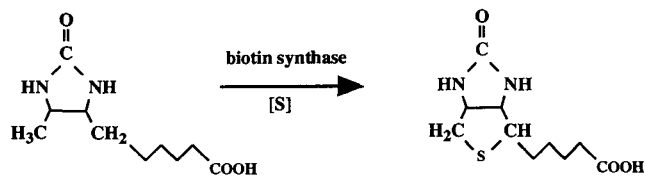


Figure 1. Reaction catalyzed by biotin synthase. Biotin synthase catalyzes the addition of a sulfur atom to dethiobiotin, converting it to biotin.

E. coli was cultured in Luria broth medium supplemented when appropriate with 50 $\mu\text{g}/\text{mL}$ ampicillin. Biotin auxotrophy of *E. coli* strains was tested on M9 minimal medium (Sambrook et al., 1989).

Nucleic Acid Manipulation

Plasmid and λ DNA for PCR, restriction enzyme analysis, and Southern analysis was isolated and purified according to standard methods (Ausubel et al., 1992). Restriction enzymes and DNA-modifying enzymes were from Promega. Template DNA for sequence analysis was purified using the Genomed JETStar miniprep kit (PGC Scientifics, Frederick, MD). *A. thaliana* DNA was isolated and purified from leaf tissue according to the method of Dellaporta et al. (1983). Site-specific mutagenesis was conducted by the method of Kunkel et al. (1987) using a Bio-Rad Mutagen kit.

Cloning

DNA primers were synthesized at the Iowa State University Nucleic Acids Facility. PCR was carried out in a total volume of 100 μL and contained 2.5 units of *Taq* polymerase, 1 μg of template DNA, 25 mM deoxyribonucleotide triphosphate, 2.5 mM MgCl_2 , 50 mM KCl, 10 mM Tris-HCl, pH 9.0, 0.1% (v/v) Triton X-100, and 30 or 300 pmol of DNA primers. PCR conditions were 1 cycle at 94°C for 3 min, 61°C for 3 min, and 72°C for 1 min; followed by 34 cycles at 94°C for 1 min, 61°C for 1 min, and 72°C for 1 min; and ending with 1 cycle at 94°C for 1 min, 61°C for 1 min, and 72°C for 10 min. PCR products were blunt ended with T4 DNA polymerase, digested with *EcoRI*, and ligated, utilizing T4 DNA ligase, into pBluescript SK(+) (Stratagene) that had been digested with *EcoRI* and *SmaI*.

The gene coding for biotin synthase was isolated from an *A. thaliana* ecotype Landsberg *erecta* genomic library (Voytas et al., 1990). This library was screened by hybridization with a ^{32}P -labeled probe made from the 86e12 clone. Isolation and purification of the genomic clone, termed λ BIO2, was performed according to standard methods (Ausubel et al., 1992). Fragments of the genomic clone were subcloned into pBluescript for sequence analysis.

Sequence Analysis

DNA sequencing was performed at the Iowa State University Nucleic Acids Facility with automated Applied Bio-

systems sequencing equipment, utilizing either T3 or T7 primers. Sequence analysis was performed using the Genetics Computer Group software package (Madison, WI).

Southern and Northern Analyses

DNA for Southern analysis was digested with various restriction enzymes and fragments were separated by electrophoresis in a 0.8% agarose gel. DNA was transferred to Magna nylon-supported nitrocellulose (Micron Separations, Westborough, MA) and bound to the filter by baking for 2 h at 80°C under vacuum. Prehybridization (5 \times SSC, 1 \times Denhardt's solution, 50 mM Tris-HCl, pH 8.0, 0.2% SDS, 10 mM EDTA, and 10 mg/mL fish sperm DNA) was for 1 h at 65°C. Hybridization was carried out for 16 to 18 h at 65°C in a solution with identical composition but with the addition of dextran sulfate to a final concentration of 10% and ^{32}P -labeled DNA probe (final concentration 5 ng/mL, 10⁹ cpm/mg DNA). DNA probes were labeled with [α - ^{32}P]dCTP using a random primer method. Following hybridization, filters were washed in 2 \times SSC at room temperature for 30 min (low stringency) or in 0.2 \times SSC, 2% SDS at 65°C for 30 min (high stringency).

RNA was isolated from tissues following extraction with an SDS/phenol-containing buffer (Logemann et al., 1987). RNA was fractionated by electrophoresis in a 1.2% (w/v) agarose gel that contained formaldehyde (Lehrach et al., 1977). The fractionated RNA was transferred to a Magna graph nylon membrane (Micron Separations) with 10 \times SSC, and the resulting blot was subjected to hybridization with a ^{32}P -labeled DNA probe coding for biotin synthase.

RESULTS AND DISCUSSION

Identification of the cDNA Coding for the Arabidopsis Biotin Synthase

The complete sequence of the EST 86e12 was determined. The 878-nucleotide cDNA contains an ORF of 696 nucleotides that codes for a protein with an amino acid sequence similar to the C-terminal two-thirds of bacterial (Otsuka et al., 1988; Sakuri et al., 1993) and yeast (Zhang et al., 1994) biotin synthases (Fig. 2). Since this ORF began at the 5' end of the cDNA and did not begin with a Met residue, it appeared that the 86e12 cDNA was not a full-length copy of the corresponding mRNA. The 5' end of the corresponding mRNA was isolated by PCR amplification from a developing silique cDNA library. This PCR was performed with 30 pmol of a primer that is complementary to a sequence 170 nucleotides downstream of the 5' end of the 86e12 cDNA (5'-GTAGTGATGACGTTTGGGTAGTAC-3') and 300 pmol of a primer complementary to the sequence flanking the *EcoRI* cloning site of the λ gtl0 vector in which the cDNA library was cloned (5'-GAGCAATTCAGCCTGGTTAAGT-3'). The anticipated PCR products overlapped the 86e12 cDNA and were detectable by Southern analysis. Analysis of the PCR products by gel electrophoresis followed by staining with ethidium bromide revealed products of between 400 and 700 nucleotides. These products hybridized to the 86e12 cDNA probe and were

<i>A. thaliana</i>	1	50
<i>E. coli</i>	MLLVRSVFRS	QLRPSVSGGL QSASCYSLSL AASAEAERTI REGPRNDWSR
<i>S. marcescens</i>		MAH-PR-TL
<i>S. cerevisiae</i>		MMAD-IH-TV
<i>B. sphaericus</i>		MPQL NRQLHPQRLV PGCSTTICIVF
		MNWL
<i>A. thaliana</i>	51	84
<i>E. coli</i>	DEIKSVYDSP LL..DLLF..HGAQV.HRH VHNFREYQQC
<i>S. marcescens</i>	SQVTELEPK- --.----EAQQ-.....-Q HFDP-Q--VS
<i>S. cerevisiae</i>	GQAQALF-K- --.E----EAQT-.....-Q HFDP-Q--VS
<i>B. sphaericus</i>	RLT--FV-KI ATRKN-SY..PT-RTY SCSHNCS- K W-DPTK-L-
	QLADE-IAGK VISD-EALAI	LNSDDDDILK LMDGAFAI-K HYYGK-KLN
<i>A. thaliana</i>		133
<i>E. coli</i>	TLLSIKTGGC SEDCSYCPQS	SRVSTGVKQA RLMSKDAVID AAKKA.KEAG
<i>S. marcescens</i>	-----A- P---K---T	---K--LE-E ---EVEQ-LE S-R--A--
<i>S. cerevisiae</i>	-----A- P---K---T	---K--LESE ---QVEQ-LE S-R--AN-
<i>B. sphaericus</i>	--MN--S-- --K-A--	--ND--L--E KMKV-E-K RGRRC-RN-
	MIMNA-S-Y- P---G--S--	--KSTAPIEKY PFIT-EEILA G--R-FENKI
<i>A. thaliana</i>		182
<i>E. coli</i>	STRFCMGAAW RDTIGRKNF	SQILEYIKEI RGM.GMEVCC TLMGIEKQQA
<i>S. marcescens</i>	----- KNPHE-DMPY	LEQM..VQGV KA--L-A-M ---TLSES--
<i>S. cerevisiae</i>	----- L-----	--MK--SAM KR-Q-MVTKV ND--L-T-V ---VDQD--
<i>B. sphaericus</i>	G-.Y-IV-SG -GPTRKDV-V	VS..-AVE-- KARY-LK--A C--LLKEE--
<i>A. thaliana</i>		232
<i>E. coli</i>	LELKKAGLTA YNHNLDTSRE	YYPNVITTRS YDDRLETLSH VRDAGINVCS
<i>S. marcescens</i>	QR-AN--DY -----	PE F-G-I---T -QE--D--EK -----K---
<i>S. cerevisiae</i>	ER-AE--DY -----	PE F-GS-I----- -QE--D--DK -----K---
<i>B. sphaericus</i>	KQ--D-----	-I----- H-SK-----T -----Q-IKN -QES--KA-T
	QQ--E--VDR ----N--ER	HHSYIT--HT -ED-VN-VEV -KKH--SP--
<i>A. thaliana</i>		278
<i>E. coli</i>	GGIIGLGEAE EDRIGLLHTL	ATLPSHPESV PINALLAVKG TPLEDO...
<i>S. marcescens</i>	--V---TV K--A---LQ-	-N--TP---- --M-VK---- --ADN....
<i>S. cerevisiae</i>	--V---TV R--A---VQ-	-N--KP---- --M-VK---- --ADN....
<i>B. sphaericus</i>	--L---S- D-H--FYI--	SNMSP--L -R-V-I-- --MAEELADP
	-A--MK-TK M-VVEIARA-	HQ--..DADSI -V-F-H-ID- -K--GT....
<i>A. thaliana</i>		326
<i>E. coli</i>	..KPVEIWEW IRMIGTARIV	MPKAMVRLSA GRVRFMSSEQ ALCFLAGANS
<i>S. marcescens</i>	..DD-DAFDF --T-AV---M	--TSY----- --EQMNEQT- -M--M-----
<i>S. cerevisiae</i>	..DD-DFDF --T-AV---M	--SSY----- --EQMNEQT- -M--M-----
<i>B. sphaericus</i>	KS-KLQFD-I L-T-A-----	---II--A- -YTMKET-- FV--M--C-
	..QDLNFRYC LKVLALP-YM	N-SKET-I-G --EVLNGLF- PFGLY-....
<i>A. thaliana</i>		376
<i>E. coli</i>	IPTGKELLTT PNDFDADQL	MFKTLGLIPK PPSFSEDDSE SENCEKVASA
<i>S. marcescens</i>	--YGC-----	--PEE-K-LQ L-RK--NPQ QTAVLAG-N- QQQRLEQALM
<i>S. cerevisiae</i>	--YGC-----	--PEE-K-LQ L-RK--NPQ QTATHEG-NQ QQQRVLAQQLL
<i>B. sphaericus</i>	--K-M--	IY-GW-E-KA -LAKW--QPM EAFKYDRS*
	--V-.DY--	EQQEANS-YR -LED--FEIE LTQKQ-EAFC*
<i>A. thaliana</i>	SH*	
<i>E. coli</i>	TPDTDEYYNA AAL*	
<i>S. marcescens</i>	NADTAEFYNA AP*	

Figure 2. Comparison of the primary structures of biotin synthases. The amino acid sequence of the *A. thaliana* biotin synthase is compared to the sequences of the enzymes from *E. coli* (Otsuka et al., 1988), *S. marcescens* (Sakuri et al., 1993), *S. cerevisiae* (Zhang et al., 1994), and *B. sphaericus* (Ohsawa et al., 1989). Residues identical to the *A. thaliana* enzyme are indicated by dashes (-), and gaps inserted to maximize alignment are indicated by dots (.). Residues 140 to 378 were deduced from the sequence of the EST 86e12.

subjected to a second round of PCR and subcloned into the pBluescript SK(+) vector.

Four cloned PCR products were isolated and sequenced. They varied in length from 380 to 680 nucleotides and only the longest clone was used in subsequent studies. The sequences of these clones extended the 86e12 cDNA sequence by 502 nucleotides to a total length of 1348 nucleotides. Northern analysis of total Arabidopsis RNA probed with the 86e12 cDNA indicates a single hybridizing RNA species of about 1.4 kb (see below). Therefore, it is likely that the sequenced cDNA represents the entire sequence of the corresponding mRNA [the entire cDNA cloned in pBluescript SK(+) is called pBS-1].

The complete sequence of pBS-1 is nearly identical to the sequences obtained by D.A. Patton, M. Pacella, and E.R.

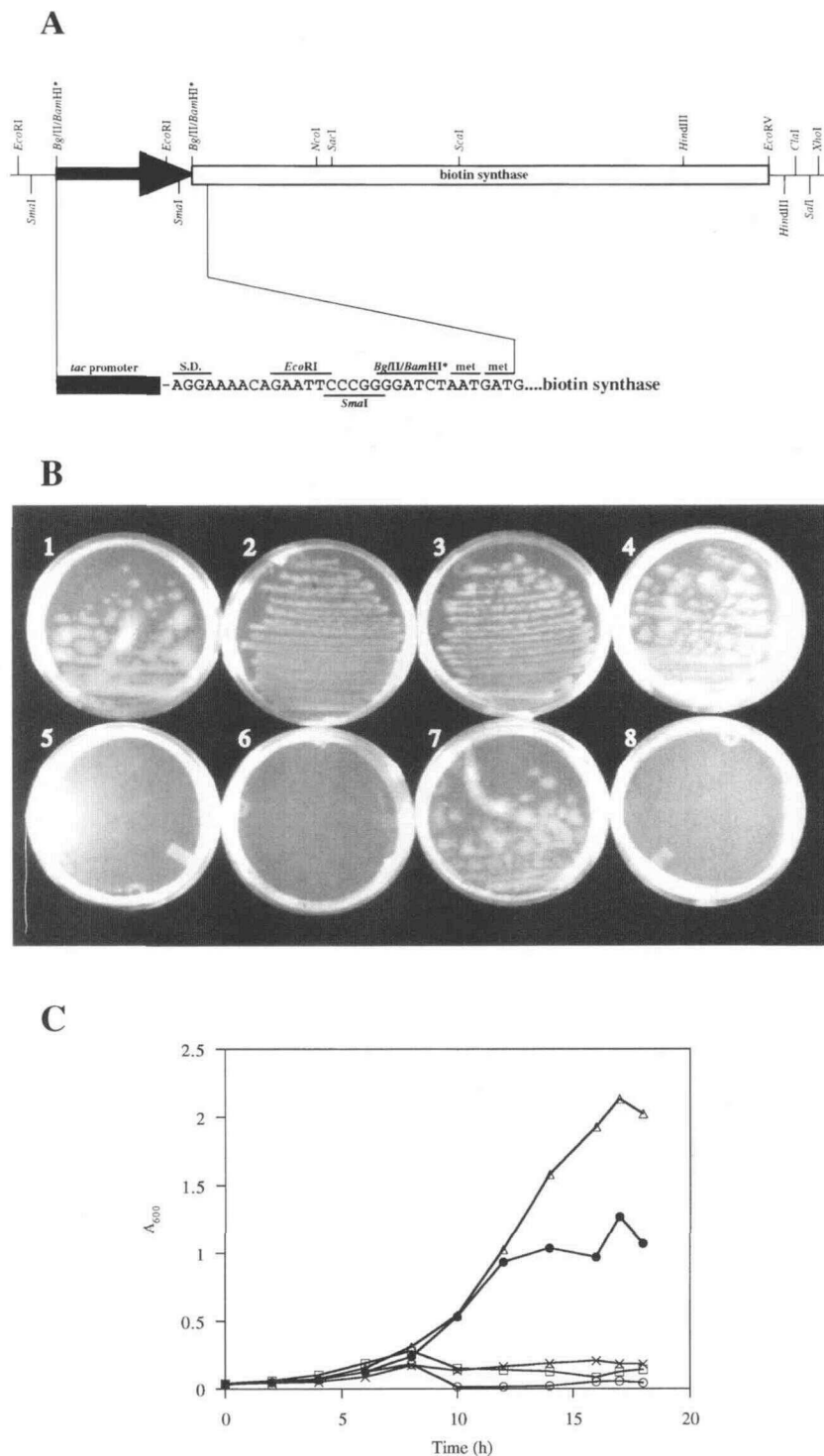
Ward (personal communication) and to the partial sequence reported in GenBank accession number L34413. The major differences between these three sequences are at the 5' and 3' noncoding ends. At the 5' end, the pBS-1 sequence is 122 nucleotides longer than the sequence shown in GenBank accession number L34413, and the 5'-most 8 nucleotides of pBS-1 are very different from those in the sequence of Patton et al. This latter difference at the 5' end of pBS-1 was confirmed upon sequencing of the corresponding gene (see below). At the 3' end, the pBS-1 sequence contains an extra 58 and 19 nucleotides upstream of the poly(A) tail that are not present in the sequence shown in GenBank accession number L34413 or in the sequence of Patton et al., respectively. In addition, pBS-1 lacks the cytosine nucleotide at position 1085 in the sequence shown in GenBank accession number L34413.

The complete sequence of pBS-1 contains an ORF that begins at the 57th nucleotide and continues for 1134 nucleotides, coding for a protein of 379 amino acids with a predicted molecular mass of 41,648 D. Comparison of the amino acid sequence of the ORF with the sequence data bases indicates that this cDNA may code for the biotin synthase of Arabidopsis (Fig. 2). Direct demonstration of the identity of pBS-1 was obtained by expressing the cDNA in the *E. coli* strain R875, which carries a mutant allele of the *bioB* gene that codes for biotin synthase. Due to this mutation, strain R875 cannot grow on media that lack biotin.

To express the putative Arabidopsis biotin synthase in *E. coli*, the 260-bp *Bam*HI fragment from pKK223-3, which contains the *tac* promoter (Brosius et al., 1985), was cloned into a unique *Bgl*III site created by site-specific mutagenesis immediately upstream of the first Met codon of the biotin synthase ORF. The resulting plasmid is called pBT-1. A schematic of the *tac*-biotin synthase gene carried on pBT-1 is shown in Figure 3A. In addition, we introduced a point mutation in the cDNA sequence of pBT-1 that changed Thr¹⁷³, which is highly conserved in biotin synthases, to an Ile residue; the resulting plasmid was called pBT-T173I. The plasmids pBluescript SK(+), pBS-1, pBT-1, and pBT-T173I were transformed into *E. coli* strain R875. The resulting strains were cultured on M9 minimal medium, with and without biotin (Fig. 3, B and C). As expected, strain R875 did not grow unless exogenous biotin was present in the medium (not shown). Furthermore, R875 harboring pBluescript SK(+), pBS-1, or pBT-T173I did not grow unless exogenous biotin was present in the medium. However, R875 harboring pBT-1 grew on minimal medium in the absence of exogenous biotin. Thus, expression of the Arabidopsis cDNA complemented the missing biotin synthase function, showing directly that the cDNA in pBS-1 codes for biotin synthase. Furthermore, since pBT-T173I was unable to complement the *bioB* mutation, Thr¹⁷³ is likely to be a residue that is important for the catalytic competence of the enzyme.

E. coli R875 containing pBT-1 was grown at both 30 and 37°C. The complementation of the *bioB* mutation by the expression of the Arabidopsis biotin synthase was

Figure 3. Complementation of the *bioB* mutant of *E. coli* by expression of the Arabidopsis biotin synthase. A, Schematic representation of the *tac*-biotin synthase cDNA in plasmid pBT-1 used to express biotin synthase in *E. coli*. The 260-bp *Bam*HI fragment from pKK223-3 (Brosius et al., 1985) containing the *tac* promoter (arrow) was cloned into a unique *Bgl*II site created by site-specific mutagenesis immediately upstream of the translation-initiating ATG codon of the biotin synthase cDNA (open rectangle). The Shine-Delgarno sequence (S.D.) is situated 23 nucleotides upstream of the translation-initiating ATG codon. The cloning of the *tac* promoter destroyed the *Bgl*II and *Bam*HI cloning sites (*Bgl*II/*Bam*HI*). B, The *E. coli* strain R875, which carries the *bioB17* allele, cannot grow on media lacking biotin (Cleary and Campbell 1972). This strain was transformed with the plasmids pBluescript SK(+) (plates 1 and 5), pBS-1 (plates 2 and 6), pBT-1 (plates 3 and 7), and pBT-T1731 (plates 4 and 8). The resulting strains were plated on M9 minimal medium containing 100 μ g/mL ampicillin and either with 10 mM biotin (plates 1-4) or without any biotin (plates 5-8) and incubated at 30°C. C, The *E. coli* strain R875 transformed with the plasmids pBluescript SK(+) (○, ●), pBS-1 (□), pBT-1 (△), and pBT-T1731 (×) were inoculated into liquid M9 minimal medium containing 100 μ g/mL ampicillin and either with 10 mM biotin (●) or without any biotin (○, □, △, ×). Growth of each culture was monitored by the increase in A_{600} .



optimal at 30°C. Complementation was achieved at 37°C; however, these cultures grew more slowly than at 30°C. Furthermore, when expression of the Arabidopsis biotin synthase from pBT-1 was induced with isopropylthio- β -galactoside, growth of the culture was slowed even further regardless of whether the medium for this experiment was Luria broth or M9 (data not shown). These data are consistent with the observation that the overex-

pression of the *E. coli bioB* gene is toxic to cells (Ifuku et al., 1995).

Primary Structure of Biotin Synthase

The deduced amino acid sequence of the *A. thaliana* biotin synthase has significant homology to biotin synthases of bacterial and yeast origins (Fig. 2; Table I). Of

Table 1. Sequence homologies between biotin synthases

The homology between pairs of biotin synthases is indicated by the percentage of identical and similar (in parentheses) amino acids in the sequences.

	<i>A. thaliana</i>	<i>E. coli</i>	<i>S. marcescens</i>	<i>S. cerevisiae</i>	<i>B. sphaericus</i>
<i>A. thaliana</i>	100 (100)	53 (69)	53 (69)	49 (66)	33 (56)
<i>E. coli</i>		100 (100)	87 (93)	46 (67)	34 (59)
<i>S. marcescens</i>			100 (100)	47 (66)	34 (61)
<i>S. cerevisiae</i>				100 (100)	32 (55)
<i>B. sphaericus</i>					100 (100)

these biotin synthases, the *E. coli* (Otsuka et al., 1988), the *S. marcescens* (Sakuri et al., 1993), and the *S. cerevisiae* (Zhang et al., 1994) enzymes are most similar in sequence to the Arabidopsis biotin synthase (approximately 50% sequence identity), whereas the *B. sphaericus* biotin synthase (Ohsawa et al., 1989) is more diverse in sequence (33% sequence identity).

The sequence similarities between the biotin synthases fall into three domains. The first of these domains, between residues Thr⁸⁵ and Trp¹⁴³, contains 30 residues that are identical in 4 of the biotin synthases, 12 of which are identical in all 5 biotin synthases. This domain contains the motif ⁹⁴Cys-X-X-X-Cys-X-X-Cys, which is conserved in all of the biotin synthases sequenced to date. Furthermore, this motif also occurs in the NifB protein of nitrogen-fixing bacteria (Joerger and Bishop, 1988) and the LipA protein of *E. coli* (Hayden et al., 1992; Reed and Cronan, 1993). The NifB protein is thought to be involved in the synthesis of the iron-molybdenum cofactor of nitrogenases. The LipA protein appears to be involved in the insertion of the first sulfur on the carbon backbone of lipoic acid, a reaction that is very similar to that catalyzed by biotin synthase. The motif Cys-X-X-X-Cys-X-X-Cys may represent a combination of two related motifs that occur in proteins that contain Fe-S clusters. These motifs are Cys-X-X-Cys and Cys-X-X-X-Cys (Stout, 1982). Indeed, biotin synthase is a [2Fe-2S] cluster protein (Sanyal et al., 1994), and the Cys residues in this motif may coordinate to the Fe atoms of the Fe-S cluster.

The second conserved domain, from residues Gly¹⁶⁷ to Leu²⁷⁵, contains 35 residues that are identical in all 5 biotin synthase sequences. This domain contains the Thr¹⁷³ that we found to be important in the reaction catalyzed by biotin synthase. One of the most highly conserved portions of this domain contains the only His residue that is conserved in all five biotin synthases. This His residue occurs in the motif Tyr¹⁹³-Asn-His-Asn-Leu/Ile; a very similar motif also occurs in the LipA protein of *E. coli*, where the sequence is Phe-Asn-His-Asn-Leu (Hayden et al., 1992; Reed and Cronan, 1993). The structural and chemical function of this His residue and the Cys residues of the Cys⁹⁴-X-X-X-Cys-X-X-Cys motif may be similar in biotin synthase and the LipA protein. This is indicated by the fact that these 2 motifs are separated by 99 residues in both biotin synthase and LipA. These motifs may be jointly involved in the binding of metal cofactors, such as the Fe of the Fe-S cluster(s).

The third conserved domain, near the carboxyl end of the protein, is between residues Ile²⁸⁷ and Leu³⁵³. This domain

contains 15 residues identical in all 5 biotin synthase sequences.

The Arabidopsis biotin synthase sequence is slightly longer than the microbial biotin synthases sequenced to date (Fig. 2). In particular, comparison of the N-terminal portion of these sequences clearly indicates that the Arabidopsis protein contains a 40-residue extension relative to the biotin synthases from prokaryotic organisms. The amino acid composition and sequence of this 40-amino acid extension have many characteristics in common with transit peptides for targeting to the mitochondria or chloroplasts (Attardi and Schatz, 1988). Namely, this sequence has an overall positive charge due to the relatively high proportion of Arg residues (4 of 40) and a lack of negatively charged residues, and is rich in hydroxylated residues (11 of 40 residues are Ser and Thr). In addition, theoretical secondary structure predictions indicate that this putative transit peptide would fold into an amphiphilic α helix or β sheet. Furthermore, the sequence Arg³⁸-Thr-Ile-Arg is similar to a motif (Arg-X- \downarrow -[Phe/Ile/Leu]-Ser) that may represent the potential cleavage site for a mitochondrial transit peptide (Heijne, 1992). Indeed, analysis of the primary sequence of the Arabidopsis biotin synthase by the PSORT algorithm (Nakai and Kanehisa, 1992) predicts that this protein is targeted to mitochondria. Although the primary sequence of the Arabidopsis biotin synthase predicts that the enzyme may be present in an organelle, possibly the mitochondrion, this will need to be shown directly, particularly in light of the fact that most of the free biotin in mesophyll cells of leaves has been shown to be in the cytosol (Baldet et al., 1993a) and that a pimeloyl-CoA synthetase that may be involved in biotin biosynthesis has been shown to be in the cytosol (Gerbling et al., 1994). If biotin synthase is located in an organelle, then organelle-synthesized biotin would need to be exported to accumulate in the cytosol.

Accumulation of the Biotin Synthase mRNA

The mRNA coding for biotin synthase was detected by northern analyses of RNA isolated from leaves and siliques of Arabidopsis plants. Hybridization of these RNA blots with the biotin synthase cDNA (Fig. 4A) revealed the approximately 1.4-kb biotin synthase mRNA. Although we did not attempt to ascertain the absolute level at which the biotin synthase mRNA accumulates, we were surprised to find that it is a readily detectable mRNA in northern blot analyses of total RNA preparations. Thus, we believe that

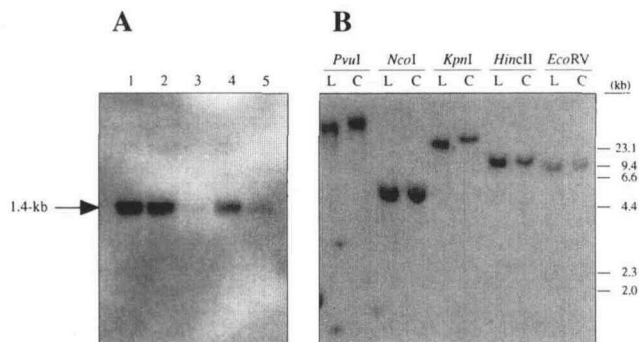


Figure 4. Biotin synthase mRNA and gene. A, Accumulation of the biotin synthase mRNA. RNA was isolated from rosette leaves of 3- (lane 1), 16- (lane 2), and 28-d-old (lane 3) *Arabidopsis* plants, and from young siliques (5–7 d after flowering; lane 4) and mature siliques (10–13 d after flowering; lane 5). Twenty micrograms of RNA were fractionated by electrophoresis in a formaldehyde-containing gel and subjected to northern blot analysis. The blot was probed with 32 P-labeled biotin synthase cDNA. B, DNA was isolated from leaves of the Columbia (C) and Landsberg *erecta* (L) ecotypes of *Arabidopsis*, digested with the indicated restriction enzymes, and subjected to Southern blot analysis. The blot was probed with the EST 86e12.

the biotin synthase mRNA may not be a low-abundance mRNA.

The accumulation of the biotin synthase mRNA was investigated in rosette leaves and siliques, and its accumulation appeared to be under developmental regulation. In both leaves and siliques, abundance of this mRNA was higher in young organs and decreased dramatically as the organ aged. Young leaves (from 3- and 16-d-old plants) accumulated higher levels of the biotin synthase mRNA than young developing siliques (5–7 d after flowering). Because biotin synthase is a member of a linear biosynthetic pathway, it will be interesting to ascertain if the accumulation of the biotin synthase mRNA is coordinated with the accumulation of mRNAs coding for other enzymes of the pathway. Currently, this question cannot be directly addressed. However, evidence has been gathered that indicates that expression of genes required for biotin biosynthesis may be linked: the accumulation of the biotin synthase mRNA is induced as the homozygous *bio1/bio1* plants deplete their biotin reserve (D.A. Patton, personal communication).

Isolation and Characterization of the Arabidopsis Biotin Synthase Gene

Southern blots of genomic DNA isolated from *Arabidopsis thaliana* ecotypes Columbia and Landsberg *erecta* were probed with the 86e12 cDNA. These analyses show a simple hybridization pattern, with a single hybridizing band in each of the restriction enzyme digests (Fig. 4B). Even under low-stringency washes ($2\times$ SSC at 25°C), only the hybridizing DNA fragments shown in Figure 4B were detectable (data not shown). These data indicate that in *Arabidopsis* biotin synthase is encoded by a single gene. We propose naming the *Arabidopsis* biotin synthase gene *BIO2*, consistent with the precedent established by *bio1*, a biotin auxo-

trophic mutant of *Arabidopsis*. The mutation in *bio1* appears to affect the second enzyme of biotin biosynthesis, 7,8-diaminopelargonic acid amino transferase (Shellhammer and Meinke, 1991; Patton et al., 1995). In *E. coli*, the *bioA* gene codes for 7,8-diaminopelargonic acid amino transferase and *bioB* codes for biotin synthase (Eisenberg, 1987). Thus, the *BIO2* gene of *Arabidopsis* is the homolog of the *bioB* gene of *E. coli*.

An *Arabidopsis*, ecotype Landsberg *erecta*, genomic library (Voytas et al., 1990) was screened by hybridization with the 86e12 cDNA clone. A single genomic clone, called λ BIO2, was isolated. Southern blot analysis of this clone probed with fragments specific for the 3' and 5' ends of the biotin synthase cDNA identified three overlapping fragments that contained the biotin synthase gene (data not shown). These three fragments, a 1.3-kb *SacI* fragment, a 1.8-kb *HindIII* fragment, and a 1.2-kb *XbaI* fragment, were subcloned into pBluescript SK(+) and sequenced.

The biotin synthase gene sequence is interrupted by five introns that are between 89 (intron 5) and 340 nucleotides (intron 1) in length (Fig. 5). The resulting six exons range in lengths from 101 (exon 3) to 456 nucleotides (exon 6). The protein coding region of the biotin synthase gene has a length of 2230 nucleotides, with the 3' end of the transcribed region of the gene at position 2388 (numbered relative to the adenosine of the first Met codon). The 5' end of the transcribed region of the biotin synthase gene has not been directly determined, although it must be at, or slightly upstream of, position -56 , which is the 5' end of the longest cDNA sequenced. A putative TATA box at position -172 to -160 has the sequence TAAAGATATATGA, which is 62% identical to the consensus plant TATA box TCACTATATAG (Joshi, 1987). Since TATA boxes usually occur between 20 and 40 nucleotides upstream of the transcript start site, the biotin synthase mRNA may begin somewhere between positions -140 and -120 .

In addition to the putative TATA box, there are four imperfect palindromes in three regions of the DNA upstream of and including the translation start site. The first palindromic sequence, ACGACGTCGCGTAGAACACTT-TGTTACGACGTTGT, is found between positions -353 and -326 . The second palindrome, from -209 to -172 , contains two sequences with dyad symmetry. One is the sequence AAGCCCAACATATAAATTTGGGCTT and the

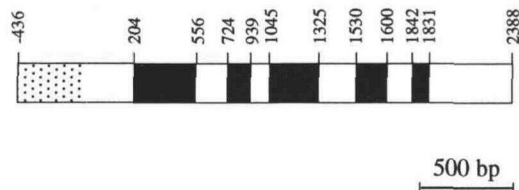


Figure 5. Schematic representation of the structure of the *Arabidopsis* *BIO2* gene, which codes for biotin synthase. A 2824-bp genomic fragment containing the coding sequence for biotin synthase was completely sequenced. The sequence is numbered relative to the adenosine nucleotide of the translation-initiating ATG codon. The sequence contains 436 bp that are upstream of the 5' end of the longest cDNA (stippled area), which represents the potential promoter. Introns are represented by black boxes.

other is AATTGGGCTTATTACAGCCCAGTT, which overlaps the first inverted repeat in this region and the first base of the putative TATA box. The third palindrome, CAGATCAAATGATGCTTGTTCGATCTG, overlaps the translation start site. This palindromic sequence, with the translation start site shown in boldface type, is transcribed and may form a stem-loop structure in the biotin synthase mRNA. The palindromic sequences in the promoter region of the biotin synthase gene may function in the regulation of its transcription.

Two unusual pyrimidine-rich sequences that contain repeated CTT and CT motifs are found in the transcribed but untranslated regions of the biotin synthase gene. One of these occurs in the 5'-transcribed but untranslated region of the gene and contains five CTT repeats ending with CCAC. It is interesting that this sequence is also found in the 5'-transcribed but untranslated region of the PEP carboxylase mRNA from the ice plant, *Mesembryanthemum crystallinum* (Cushman and Bohnert, 1992). A second pyrimidine-rich sequence, consisting of 4 CTT motifs followed by 24 CT motifs, occurs in the first intron. Such polypyrimidine tracts of CTT and CT repeats have been correlated with nonB DNA structures (Wells et al., 1988).

The isolation of the *BIO2* gene provides molecular tools with which the regulation of biotin biosynthesis can be studied. These studies should reveal how biotin becomes available for biotinylation of biotin-containing enzymes that are compartmentalized in different subcellular compartments. For example, methylcrotonyl-CoA carboxylase is in the mitochondria (Baldet et al., 1992; Song et al., 1994), and acetyl-CoA carboxylase isozymes are present in chloroplasts (Choi et al., 1995) and cytosol (Alban et al., 1994; Shorosh et al., 1994; Yanai et al., 1995). In addition, such studies should reveal the mechanism(s) by which plants accumulate most of their biotin as protein-free biotin (Shellhammer and Meinke, 1990; Baldet et al., 1993a; Wang et al., 1995), whereas bacteria use all of their synthesized biotin for biotinylation of biotin-containing proteins (Cronan, 1989).

ACKNOWLEDGMENTS

We thank Dr. David A. Patton and colleagues (Ciba-Geigy Corporation, Research Triangle Park, NC) for sharing unpublished data.

Received August 31, 1995; accepted December 20, 1995.
Copyright Clearance Center: 0032-0889/96/110/1021/08.
The accession number for the sequence reported in this article (the *BIO2* gene) is U24147.

LITERATURE CITED

- Alban C, Baldet P, Douce R (1994) Localization and characterization of two structurally different forms of acetyl-CoA carboxylase in young pea leaves, of which one is sensitive to aryloxyphenoxypropionate herbicides. *Biochem J* **300**: 557-565
- Attardi G, Schatz G (1988) Biogenesis of mitochondria. *Annu Rev Cell Biol* **4**: 289-333
- Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1992) Short Protocols in Molecular Biology. Greene Publishing Associated/John Wiley & Sons, New York
- Baldet P, Alban C, Axiotis S, Douce R (1993a) Localization of free and bound biotin in cells from green pea leaves. *Arch Biochem Biophys* **303**: 67-73
- Baldet P, Alban C, Stella A, Douce R (1992) Characterization of biotin and 3-methylcrotonyl-coenzyme A carboxylase in higher plant mitochondria. *Plant Physiol* **99**: 450-455
- Baldet P, Gerbling H, Axiotis S, Douce R (1993b) Biotin biosynthesis in higher plant cells: identification of intermediates. *Eur J Biochem* **217**: 479-485
- Birch OM, Fuhrmann M, Shaw NM (1995) Biotin synthase from *Escherichia coli*, an investigation of the low molecular weight and protein components required for activity *in vitro*. *J Biol Chem* **270**: 19158-19165
- Brosius J, Erfle M, Storella J (1985) Spacing of the -10 and -35 regions in the tac promoter: effect on its *in vivo* activity. *J Biol Chem* **260**: 3539-3541
- Choi J-K, Yu F, Wurtele ES, Nikolau BJ (1995) Molecular cloning and characterization of the cDNA coding for the biotin-containing subunit of the chloroplastic acetyl-CoA carboxylase. *Plant Physiol* **109**: 619-625
- Cleary PP, Campbell A (1972) Deletion and complementation analysis of the biotin gene cluster of *Escherichia coli*. *J Bacteriol* **112**: 830-839
- Cronan JE (1989) The *E. coli* bio operon: transcriptional repression by an essential protein modification enzyme. *Cell* **58**: 427-429
- Cushman JC, Bohnert HJ (1992) Salt stress alters A/T-rich DNA-binding factor interactions within the phosphoenolpyruvate carboxylase promoter from *Mesembryanthemum crystallinum*. *Plant Mol Biol* **20**: 411-424
- Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA miniprep preparation version II. *Plant Mol Biol Rep* **1**: 19-21
- Eisenberg M (1987) Biosynthesis of biotin and lipoic acid. In FC Neidhardt, JL Ingraham, KB Low, B Magasanik, M Schaechter, HE Umbarger, eds, *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology. American Society for Microbiology, Washington, DC, pp 544-550
- Gerbling H, Axiotis S, Douce R (1994) A new acyl-CoA synthetase, located in higher plant cytosol. *J Plant Physiol* **143**: 561-564
- Hayden MA, Huang I, Bussiere DE, Ashley GW (1992) The biosynthesis of lipoic acid. *J Biol Chem* **267**: 9512-9515
- Heijne GV (1992) Cleavage-site motifs in protein targeting sequences. *Genet Eng* **14**: 1-11
- Ifuku O, Kishimoto J, Haze S-I, Yangi M, Fukushima S (1992) Conversion of dethiobiotin to biotin in cell-free extracts of *Escherichia coli*. *Biosci Biotechnol Biochem* **56**: 1780-1785
- Ifuku O, Koga N, Haze S-I, Kishimoto J, Arai T, Wachi Y (1995) Molecular analysis of growth inhibition caused by overexpression of the biotin operon in *Escherichia coli*. *Biosci Biotechnol Biochem* **59**: 184-189
- Joerger RD, Bishop PE (1988) Nucleotide sequence and genetic analysis of the *nifB-nifQ* region from *Azotobacter vinelandii*. *J Bacteriol* **170**: 1475-1487
- Joshi CP (1987) An inspection of the domain between putative TATA box and the translation start site in 79 plant genes. *Nucleic Acids Res* **15**: 6643-6653
- Knowles J (1989) The mechanism of biotin-dependent enzymes. *Annu Rev Biochem* **58**: 195-221
- Kunkel TA, Roberts JD, Zakur RA (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol* **154**: 367-382
- Lehrach H, Diamond D, Wozney JM, Boedtker H (1977) RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. *Biochemistry* **16**: 4743-4751
- Logemann J, Schell J, Willmitzer L (1987) Improved method for the isolation of RNA from plant tissues. *Anal Biochem* **163**: 16-20
- Nakai K, Kanehisa M (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics* **14**: 897-911
- Newman T, de Bruijn FJ, Green P, Keegstra K, Kende H, McIntosh L, Ohlrogge J, Raikhel N, Somerville S, Thomashow M,

- Retzel E, Somerville C** (1994) Genes galore. A summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. *Plant Physiol* **106**: 1241–1255
- Ohsawa I, Speck K, Kisou T, Hayakawa K, Zinsius M, Gloeckler R, Lemoine Y, Kamogawa K** (1989) Cloning of the biotin synthetase gene from *Bacillus sphaericus* and expression in *Escherichia coli* and *Bacilli*. *Gene* **80**: 39–48
- Otsuka AJ, Buoncristiani MR, Howard PK, Flamm J, Johnson C, Yamamoto R, Uchida K, Cook C, Ruppert J, Matsuzaki J** (1988) The *Escherichia coli* biotin biosynthetic enzyme sequences predicted from the nucleotide sequence of the *bio* operon. *J Biol Chem* **263**: 19577–19585
- Patton DA, Volrath S, Ward ER** (1995) Complementation of the *bio1 Arabidopsis* biotin auxotroph with a bacterial biotin biosynthetic gene. In 6th International Conference on *Arabidopsis* Research. University of Wisconsin, Madison, S237
- Reed KE, Cronan JE** (1993) Lipoic acid metabolism in *Escherichia coli*: sequencing and functional characterization of the *lipA* and *lipB* genes. *J Bacteriol* **175**: 1325–1336
- Sakuri N, Imaj Y, Masuda M, Komatsubara S, Tosa T** (1993) Molecular breeding of a biotin-hyperproducing *Serratia marcescens* strain. *Appl Environ Microbiol* **59**: 3225–3232
- Sambrook J, Fritsch EF, Maniatis T** (1989) *Molecular Cloning: A Laboratory Manual*, Ed 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Sanyal I, Cohen G, Flint DH** (1994) Biotin synthase: purification, characterization as a [2Fe-2S] cluster protein, and in vitro activity of the *Escherichia coli bioB* gene product. *Biochemistry* **33**: 3625–3631
- Schneider T, Dinkins R, Robinson K, Shellhammer J, Meinke DW** (1989) An embryo-lethal mutant of *Arabidopsis thaliana* is a biotin auxotroph. *Dev Biol* **131**: 161–167
- Shellhammer J, Meinke D** (1990) Arrested embryos from the *bio1* auxotroph of *Arabidopsis thaliana* contains reduced levels of biotin. *Plant Physiol* **93**: 1162–1167
- Shorrosh BS, Dixon RA, Ohlrogge JB** (1994) Molecular cloning, characterization, and elicitation of acetyl-CoA carboxylase from alfalfa. *Proc Natl Acad Sci USA* **91**: 4323–4327
- Song J, Wurtele ES, Nikolau BJ** (1994) Molecular cloning and characterization of the cDNA coding for the biotin-containing subunit of 3-methylcrotonyl-CoA carboxylase: identification of the biotin carboxylase and biotin-carrier domains. *Proc Natl Acad Sci USA* **91**: 5779–5783
- Stout CD** (1982) Iron-sulfur protein crystallography. In TG Spiro, ed, *Iron Sulfur Proteins*. Wiley, New York, pp 97–146
- Voytas DF, Konieczny A, Cummings MP, Ausubel FM** (1990) The structure, distribution and evolution of the *Ta1* retrotransposable element family of *Arabidopsis thaliana*. *Genetics* **126**: 713–721
- Wang X, Wurtele ES, Nikolau BJ** (1995) Regulation of β -methylcrotonyl-CoA carboxylase activity by biotinylation of the apoenzyme. *Plant Physiol* **108**: 1133–1139
- Wells RD, Collier DA, Hanvey JC, Shimizu M, Wohlrab F** (1988) The chemistry and biology of unusual DNA structures adopted by oligopurine-oligopyrimidine sequences. *FASEB J* **2**: 2939–2949
- Yanai Y, Kawasaki T, Shimada H, Wurtele ES, Nikolau BJ, Ichikawa N** (1995) Genetic organization of the 251 kDa acetyl-CoA carboxylase genes in *Arabidopsis*: tandem gene duplication has made two differentially expressed isozymes. *Plant Cell Physiol* **36**: 779–787
- Zhang S, Sanyal I, Bulboaca GH, Rich A, Flint D** (1994) The gene for biotin synthase from *Saccharomyces cerevisiae*: cloning, sequencing and complementation of *Escherichia coli* strains lacking biotin synthase. *Arch Biochem Biophys* **309**: 29–35