Structure of a gene encoding a cytosolic acetyl-CoA carboxylase of hexaploid wheat

J. Podkowinski*, G. E. Sroga, R. Haselkorn, and P. Gornick†

Department of Molecular Genetics and Cell Biology, University of Chicago, 920 East 58th Street, Chicago, IL 60637

Contributed by R. Haselkorn, November 17, 1995

ABSTRACT An entire gene encoding wheat (var. Hard Red Winter Tam 107) acetyl-CoA carboxylase [ACCase; acetyl-CoA:carbon-dioxide ligase (ADP-forming), EC 6.4.1.2] has been cloned and sequenced. Comparison of the 12-kb genomic sequence with the 7.4-kb cDNA sequence reported previously revealed 29 introns. Within the coding region, the exon sequence is 98% identical to the known wheat cDNA sequence. A second ACCase gene was identified by sequencing fragments of genomic clones that include the first two exons and the first intron. Additional transcripts were detected by 5′ and 3′ RACE analysis (rapid amplification of cDNA ends). One set of transcripts had a 5′ end sequence identical to the cDNA found previously and another set was identical to the gene reported here. The 3′ RACE clones fall into four distinguishable sequence sets, bringing the number of ACCase sequences to six. None of these cDNA or genomic clones encodes a chloroplast targeting signal. Identification of six different sequences suggests that either the cytosolic ACCase genes are duplicated in the three chromosome sets in hexaploid wheat or that each of the six alleles of the cytosolic ACCase gene has a readily distinguishable DNA sequence.

Acetyl-CoA carboxylase [ACCase; acetyl-CoA:carbon-dioxide ligase (ADP-forming), EC 6.4.1.2] catalyzes the first committed step in fatty acid biosynthesis and provides malonyl-CoA for the synthesis of a variety of important secondary metabolites and for malonylation. In plants, these primary and secondary metabolic pathways are located in different compartments (plastids and cytosol) accounting for the existence of multiple isozymes bearing eukaryotic or prokaryotic features. It has been shown recently that plants have indeed more than one form of ACCase (reviewed in ref 1). The one located in plastids, the primary site of plant fatty acid synthesis, can be either a eukaryotic-type high molecular weight multifunctional enzyme (e.g., in wheat and maize) or a prokaryotic-type multisubunit enzyme (e.g., in pea, soybean, tobacco, and Arabidopsis). The other plant ACCase, located in the cytosol, is of the eukaryotic type.

In Gramineae, genes for both cytosolic and plastid eukaryotic-type ACCases are nuclear. No ACCase coding sequence can be found in the complete sequence of rice chloroplast DNA. The plastid ACCase in maize (2) and cytosolic ACCase in wheat, reported here, are nuclear encoded. In other plants, subunits of ACCase other than the carboxyltransferase subunit encoded by a homolog of the Escherichia coli accD gene, present in the chloroplast genome (1, 3), must be also encoded in the nuclear DNA. Like the vast majority of plastid proteins, plastid ACCases are synthesized in the cytoplasm and then transported into the plastid. The amino acid sequence of the cytosolic and some subunits of the plastid ACCases from several plants have been deduced from genomic or cDNA sequences (2–9).

There is experimental evidence suggesting that, in plants, ACCase activity controls carbon flow through the fatty acid pathway and therefore may serve as an important regulation point of plant metabolism (10–12). These studies must take into account the existence of isozymes located in different cell compartments, regulated independently and with different structural and biochemical properties, including different sensitivity to herbicides. Previously, we described the cloning and sequencing of cDNA for wheat ACCase (4). In this paper, we report the structure of the corresponding gene. 3

MATERIALS AND METHODS

Isolation and Analysis of ACCase Genomic Clones. A wheat genomic library (Triticum aestivum, var. Hard Red Winter Tam 107, 13-day light-grown seedlings) was purchased from Clontech. This λEMBO3 library was prepared from genomic DNA partially digested with Sau3A. Colony ScreenPlus (DuPont) membrane was used according to the manufacturers’ protocol (hybridization at 65°C in 1 M NaCl/10% dextran sulfate). The library was screened with a 440-bp PCR-amplified fragment of ACCase-specific cDNA and with cDNA clone 24-3 (4). In each case, ≈1.2 × 106 plaques were tested. Twenty-four clones containing ACCase-specific DNA fragments were purified and mapped. Selected restriction fragments of these genomic clones were subcloned into pBluescript SK for further analysis and cloning. The 5′-terminal fragment of the gene (clone 145) was amplified by PCR using wheat genomic DNA as a template. Primers were based on the sequence of genomic clone 233 (CGCTATAGGGAAAAAGTATAGGATGTTGG) and 3′ RACE (rapid amplification of cDNA ends) clone 4 (ATCGATGCGCCTGGCTCCTCAATTTCTATT). All PCR components except Taq polymerase were incubated for 5 min at 95°C. The reactions were initiated by the addition of the polymerase followed by 35 cycles of incubation at 94°C for 1 min, 55°C for 2 min, and 72°C for 2 min. A 1.8-kb PCR product was gel-purified, reamplified using the same primers, cloned into the Invitrogen vector pCRII, and sequenced.

Analysis of mRNA by RACE. Two sets of 15 and 20 cDNA fragments corresponding to mRNA 5′ and 3′ ends, respectively, were prepared by T/A cloning of RACE products into the vector pCRII. Total RNA from 15-day-old wheat (T. aestivum var. Tam 107, Hard Red Winter) plants was prepared as described (13). A Gibco/BRL 5′ RACE kit was used according to the manufacturers’ protocol. For the 5′-end amplification, the first strand of cDNA was prepared with a gene-specific primer (GTTCCTTAGAAGGTCTCCAAGG) followed by addition of a homopolymeric dA-tail. DT-Anchor primer (GGCGGATCTGAGTCGAAGCT2) and a gene-specific primer (ACGGTGTGACATGATGGTGCCGAT)-

Abbreviations: ACCase, acetyl-CoA carboxylase; RACE, rapid amplification of cDNA ends.
*Present address: Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowski 12/14, 67 704 Poznan, Poland.
†To whom reprint requests should be addressed.
‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. U39321).

1870
GCTGCGCATG) were used in the first round of PCR. Universal primer (GCCGACTCGAGTCGACAAGC) and another gene-specific primer (ACCGGCTGACCATCCCATTGTGGGCAACC) were used for reamplification. The gene-specific primers were targeted to a stretch of 5'-end coding sequence identical in clones 39 and 71 that were available. Clone 71 was isolated from a AgtI cDNA library as described before using a fragment of cDNA 39 as probe (4). The same dT-anchor primer and universal primer together with a gene specific primer (GACTCATGAGATCAAGTTC) were used for first-strand cDNA synthesis and 3'-end amplification. The latter primer was targeted to the 3' end of the ACCase open reading frame.

All cloning, DNA manipulations, and gel electrophoresis were as described (14). DNA was sequenced on both strands by the dideoxynucleotide chain-termination method using dATP[35S] with Sequenase (United States Biochemical) or using the Perkin–Elmer/Applied Biosystems Taq DyeDeoxy terminator cycle sequencing kit and an Applied Biosystems model 373A DNA sequencer.

RESULTS

Analysis of Wheat Cytosolic ACCase Genes. Two cDNA fragments, one encoding a part of the biotin carboxylase domain of wheat ACCase and the other encoding a part of the carboxyltransferase, were used to isolate a set of overlapping DNA fragments covering the entire ACCase gene. Some of these genomic fragments were sequenced as indicated in Fig. 1. Where they overlap, the nucleotide sequences of clones 31, 191, and 233 are identical. These obviously derive from the same gene. cDNA clone 71 (see below) represents the transcription product of this gene (430-nucleotide identical sequence). The sequence of clone 145 obtained by PCR to cover the remaining 3'-end part of the gene differs from clone 233 by 5 of 400 nucleotides of the overlap located within the long exon 29 (Fig. 1). It must therefore derive from a different copy of the ACCase gene. The 3' RACE clone 4 (3'-4; see below) differs at 6 of 490 nucleotides in the overlap. The gene sequence is a composite of these three very closely related sequences. Its 5' end corresponds to the 5' end of clone 71 and the 3' end corresponds to the poly(A) attachment site of the 3' RACE clone 4. We assumed that no additional introns are present at the very end of the gene.

Comparison of the genomic sequence with the cDNA sequence reported previously (4) revealed 29 introns. Introns location is conserved among all three known plant ACCase genes except for two introns not present in wheat but found in rape (5), Arabidopsis thaliana (8) and soybean (9) (Fig. 1). The nucleotide sequence at splice sites fits well with the consensus for monocot plants. The A+T content of the gene exons and introns is 52% and 63%, respectively, compared to 42% and 61% found for other monocot plant genes (15). The exon coding sequence is 98% identical to that of the cDNA sequence reported earlier. This is the same degree of identity as found previously for different transcripts of the cytosolic ACCase genes in hexaploid wheat (4). The 11-amino acid sequence obtained previously for a CNBr-generated internal fragment of purified 220-kDa wheat germ ACCase (16) differs from the sequence encoded by our cDNA and genomic clones at one position, but it is identical with the corresponding cDNA sequence of the plastid ACCase from maize (2), excluding one amino acid, which could not be assigned unambiguously in our sequence.

Two additional genomic clones, 153 and 231, were also partially sequenced (Fig. 1). The sequenced fragments include parts of the first two exons and the first intron. Although cDNA corresponding exactly to genomic clone 153 is not available, the boundaries of the first intron could easily be identified by sequence comparison with cDNA clone 71 (corresponding to genomic clone 31). Clone 153 encodes a polypeptide that differs by only 1 of the first 110 amino acids of the ACCase open reading frame. The sequence of the 5' leader is also well conserved (Fig. 2), but the 5' part of the first intron of clone 153 is significantly different from that of genomic clone 31 (data not shown).

On the other hand, only the 3' splice site of an intron could be identified by sequence comparison in this part of clone 231. The sequence immediately upstream of the 3' splice site and that of the following exon is identical to that of clone 31. No sequence related to that found upstream of the first intron of clone 191 could be identified in clone 231 by hybridization.

Fig. 1. Structure of the cytosolic ACCase gene from wheat. Arrows indicate fragments of genomic clones analyzed in more detail. Sequenced fragments are marked in black. Localization of the ACCase functional domains was established by amino acid sequence comparison with other biotin-dependent carboxylases (4). BC, biotin carboxylase; BCC, biotin carboxyl carrier; CT, carboxyltransferase.
Fig. 2. Alignment of cDNA sequences corresponding to the 5' end of the mRNA encoding wheat cytosolic ACCase. cDNA sequences are shown in capital letters and the genomic sequence of clone 153, from which the putative sequence of the first intron was omitted, is shown in lowercase letters. The ACCase initiation codon and the beginning of its open reading frame are indicated by a horizontal arrow. AUG codons in the 5' leader sequence are underlined. Stop codons near the ACCase initiation codon are marked with triangles. The site of the first intron is marked with a vertical arrow. Asterisks indicate identical nucleotides. Thirteen additional 5' RACE clones were sequenced; these matched either 71L or 39L.

(including an ~6-kb fragment upstream of the ACCase open reading frame; data not shown) or by sequencing (~2 kb of the upstream fragment; data not shown). It is possible that the first intron in this gene is much larger (additional upstream introns cannot be excluded) or that the upstream exon(s) and untranscribed part of the gene has a completely different sequence. A cloning artifact cannot be ruled out. Indeed, clone 31 contained such an unrelated sequence at its 5' end (probably a ligation artifact).

Identification of three additional genomic clones with a sequence closely related to the other ACCase genes but containing no introns at several tested locations suggests the existence of a pseudogene in wheat. A fragment of clone 232 that was sequenced (data not shown) is represented in the diagram shown in Fig. 1. It is 93% and 96% identical with clone 233 at the nucleotide and amino acid levels, respectively.

Analysis of mRNA Ends. In the original library screen (4), we were unable to isolate any cDNA clones corresponding to the very ends of the ACCase mRNA. With the new sequence available it became possible to generate the missing pieces by RACE. Two sets of 5'-end RACE clones, 71L and 39L, were identified. Their sequence is identical to the sequence of cDNA clones 71 (this work) and 39 (4), respectively. The alignment of the longest sequence of each type is shown in Fig. 2. The two sequences extend from 233 to 312 nucleotides upstream of the ACCase initiation codon and define an approximate position of the transcription start site. None of our genomic clones corresponds to 39L. The presence of the first intron in the corresponding gene, therefore, could not be confirmed. For comparison, the corresponding exon sequence derived from clone 153 is also included in the alignment of Fig. 2. All three coding sequences are very similar (they differ by only one 3-amino acid deletion or one E to D substitution found within the first 110 amino acids) and none of them encodes additional amino acids at the N terminus—i.e., none of them encodes a potential chloroplast transit peptide.

The sequences of the 5' leaders differ significantly although they share some distinctive structural features. They are relatively long (at least 239–312 nucleotides as indicated by the lengths of 39L and 71L, respectively), G+C rich (67%), and contain upstream AUG codons. The open reading frames found in the leaders are 70–90 amino acids long and they end within a few nucleotides of the ACCase initiation codon. A similar arrangement was found in the sequence of genomic clone 153. The three upstream AUG codons are conserved and the presence of deletions, most of which are a multiple of 3 nucleotides, suggests at least some conservation of the open reading frames at the amino acid level. This arrangement, found in the cytosolic ACCase genes, contrasts with the majority of 5' untranslated leaders found in plants. Although much longer leader sequences containing upstream AUG codons have been reported in plants (e.g., see ref. 17), they are rare. In most cases, the first AUG codon is the site of initiation of translation of the major gene product. The upstream AUGs are believed to affect the efficiency of mRNA translation and as such may be important in the regulation of expression of some genes (18, 19). They are often found in mRNAs encoding transcription factors, growth factors, and receptors, all important regulatory proteins (20). They are also found in some plant mRNAs encoding heat shock proteins (21). The ~800-nucleotide leader intron found in both genes (clones 153 and 191) may also be important for the level and pattern of gene expression (e.g., see ref. 22).

Four different sequences and two different polyadenylation sites ~300 and ~500 nucleotides downstream of the translation stop codon, respectively, were detected among the 3' end RACE clones (Fig. 3). The sequence of the cDNA reported previously (4) and the sequence of genomic clone 145 are also different in this region, bringing the total number of different sequences to six. Differences of 3–14 nucleotides were found in pairwise comparisons among these six sequences within two stretches that include 282 nucleotides at the 5' end of the 3' RACE clones (data not shown) and 204 nucleotides at the 3' end (Fig. 3).

**DISCUSSION**

A gene encoding eukaryotic-type cytosolic ACCase from wheat, very similar in sequence to the cDNA reported (4), was cloned and sequenced. Nucleotide identity between the cDNA and the gene within the coding sequence is 98%. The putative translation start codon was assigned in the original cDNA sequence to the first methionine of the open reading frame. An in-frame stop codon is present 21 nucleotides upstream from this AUG and the conserved core of the biotin carboxylase domain begins ~20 amino acids further downstream. The gene encodes a 2260-amino acid protein with a calculated molecular mass of 252 kDa. The wheat cDNA did not encode an obvious chloroplast targeting sequence. The same is true for all the cDNA and genomic sequences described in this paper. The
GATAGTTACACACAAAGCATATGTATGCTGGATAGATATTCGGTGTGAGTTGTTGCAATGCAAGATTCATCATCTTAATTTA(A)n
GATAGTTACACACAAAGCATATGTATGCTGGATAGATATTCGGTGTGAGTTGTTGCAATGCAAGATTCATCATCTTAATTTACG(A)n
GATAGTTACACACAAAGCATATGTATGCTGGATAGATATTCGGTGTGAGTTGTTGCAATGCAAGATTCATCATCTTAATTTACGAGATACGATGTGATGATCGGTCGATGTGGTAGTTGT

3. Alignment of the wheat ACCase for which additional 3' sequence encode biotin-dependent carboxylases. ACCases from dicot and maize plastid ACCases are shown. The putative polyadenylylation signals are underlined. Asterisks indicate identical nucleotides. Sixteen additional 3' RACE clones were sequenced; these matched one or another of the four sequences shown.

cDNA for maize plastid ACCase, reported recently (2), does encode a chloroplast transit peptide.

Comparison of the ACCase sequence encoded by the gene reported in this paper with the sequence of the wheat ACCase reported previously (4) and with other representative biotin-dependent carboxylases is shown in Table 1. Wheat ACCase is most similar to other eukaryotic-type plant ACCases. Identity with other eukaryotic carboxylases is also significant. The core sequence of the most conserved ACCase domain, biotin carboxylase, is well conserved in both eukaryotic and prokaryotic biotin-dependent carboxylases. The other functional domains are less conserved (4). Among plant eukaryotic-type ACCases, the wheat cytosolic ACCase is no more similar to the maize plastid ACCase (both monocots) than it is to cytosolic ACCases from dicot plants. Clearly, cytosolic and plastid eukaryotic-type ACCases are quite distinct proteins. Another wheat ACCase for which partial sequence is available (32) is most likely a plastid isozyme. It is more similar to the maize plastid ACCase than to the wheat cytosolic enzyme. The plant prokaryotic-type plastid enzyme is more similar to bacterial, most notably cyanobacterial ACCases and to biotin-dependent carboxylases found in mitochondria, than to any of the plant cytosolic ACCases.

Sequence comparison of fragments of cDNA and genomic clones from the 3' end of the gene brings the total number of different genes encoding cytosolic ACCase in wheat to six, indicating that in hexaploid wheat there are at least two distinguishable coding sequences for the cytosolic ACCase in each of the three ancestral chromosome sets. Those two sequences might correspond to the alleles of the ACCase gene present in each ancestral chromosome set. On the other hand, it is possible that each pair of alleles has identical sequences, since the bread wheat studied is extensively inbred. If that is the case, then one or more ancestral genes has been duplicated.

Table 1. Amino acid identities between T. aestivum cytosolic ACCase and some other representative biotin-dependent carboxylases

<table>
<thead>
<tr>
<th></th>
<th>Full length</th>
<th>Biotin carboxylase domain</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Eukaryotic-type</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Triticum aestivum</em></td>
<td>Cytosolic</td>
<td>99</td>
<td>4</td>
</tr>
<tr>
<td><em>Arabidopsis thaliana</em></td>
<td>Cytosolic</td>
<td>72</td>
<td>8</td>
</tr>
<tr>
<td><em>Medicago sativa</em></td>
<td>Cytosolic</td>
<td>73</td>
<td>6</td>
</tr>
<tr>
<td><em>Brassica napus</em></td>
<td>Cytosolic</td>
<td>68</td>
<td>5</td>
</tr>
<tr>
<td><em>Zea mays</em></td>
<td>Plastid</td>
<td>71</td>
<td>2</td>
</tr>
<tr>
<td><em>Rattus rattus</em></td>
<td>Cytosolic</td>
<td>40</td>
<td>23</td>
</tr>
<tr>
<td><em>Coccolithus cerevisiae</em></td>
<td>Cytosolic</td>
<td>38</td>
<td>24</td>
</tr>
<tr>
<td><strong>Prokaryotic-type</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>Bacterial</td>
<td>—</td>
<td>33</td>
</tr>
<tr>
<td><em>Anaebaena 7120</em></td>
<td>Bacterial</td>
<td>—</td>
<td>34</td>
</tr>
<tr>
<td><em>Mycoococcus leprae</em></td>
<td>Bacterial</td>
<td>—</td>
<td>32</td>
</tr>
<tr>
<td><em>Nicotiana tabacum</em></td>
<td>Plastid</td>
<td>—</td>
<td>32</td>
</tr>
<tr>
<td><em>Rattus rattus</em> PCC</td>
<td>Mitochondrial</td>
<td>—</td>
<td>34</td>
</tr>
<tr>
<td><em>Saccharomyces cerevisiae</em> PC</td>
<td>Mitochondrial</td>
<td>—</td>
<td>32</td>
</tr>
<tr>
<td><em>Arabidopsis thaliana</em> MCCase</td>
<td>Mitochondrial</td>
<td>—</td>
<td>34</td>
</tr>
</tbody>
</table>

*Sequence deduced from cDNA sequence reported previously (product of a different allele or gene).
†Cellular localization uncertain.
‡Biotin carboxylase subunit of ACCase.
§Biotin carboxylase–biotin carboxyl carrier subunit of ACCase.
**Biotin carboxylase–biotin carboxyl carrier subunit a of propionyl-CoA carboxylase.
Much further work on the physical location of the ACCase genes reported here is needed before these possibilities can be evaluated.

The products of two different ACCase genes have been identified in human and rat (33, 34). The presence of multiple cytosolic ACCase genes in both plants (35) and mammals may reflect the need for differential expression of the enzyme in response to different environmental or developmental cues. The structure and activity of the ACCase promoters as well as the possible role of the 5' untranslated leader and the leader intron in gene expression remain to be investigated. To add complexity, a different set of genes is required in wheat to encode the large plastid ACCase. These genes have yet to be characterized.

The authors are grateful to E. Ward for advice and to J. DiMaio (CIBA-Geigy) and F. Roche (University of Chicago) for DNA sequencing. This work was supported by grants from the Department of Agriculture through the North Central Biotechnical Program and from CIBA-Geigy.