Post-translational modifications in the large subunit of ribulose bisphosphate carboxylase/oxygenase
(N-terminal acetylation/trimethyllysine/fast-atom bombardment.mass spectrometry)

ROBERT L. HOUTZ*, JOHN T. STULTS†, R. MICHAEL MULLIGAN‡, and N. E. TOLBERT§

ABSTRACT Two adjacent N-terminal tryptic peptides of the large subunit of ribulose bisphosphate carboxylase/oxygenase [3-phospho-D-glycerate carboxylase (dimmerizing), EC 4.1.1.39] from spinach, wheat, tobacco, and muskmelon were removed by limited tryptic proteolysis. Characterization by peptide sequencing, amino acid composition, and tandem mass spectrometry revealed that the N-terminal residue from the large subunit of the enzyme from each plant species was acetylated proline. The sequence of the penultimate N-terminal tryptic peptide from the large subunit of the spinach and wheat enzyme was consistent with previous primary structure determinations. However, the penultimate N-terminal peptide from the large subunit of both the tobacco and muskmelon enzymes, while identical, differed from the corresponding peptide from spinach and wheat by containing a trimethylsilyl residue at position 14. Thus, tryptic proteolysis occurred at lysine-18 rather than lysine-14 as with the spinach and wheat enzymes. A comparison of the DNA sequences for the large subunit of ribulose bisphosphate carboxylase/oxygenase indicates that the N terminus has been post-translationally processed by removal of methionine-1 and serine-2 followed by acetylation of proline-3. In addition, for the enzyme from tobacco and muskmelon a third post-translational modification occurs at lysine-14 in the form of N\(^\text{3}\)-trimethylation.

Ribulose bisphosphate carboxylase/oxygenase [RbuP\(_2\) carboxylase, 3-phospho-D-glycerate carboxylase (dimmerizing), EC 4.1.1.39] is a hexadecameric plant protein with eight large subunits (LSs) and eight small subunits (SSs) (for a recent review, see ref. 1). The LS is known to contain the active site, while the function of the SS remains obscure. The SS is nuclear-encoded and translated with a N-terminal transit sequence that targets it for import into the chloroplast. The SS is post-translationally processed before assembly with the chloroplast-encoded LS (2). Ambiguity exists with regard to post-translational processing of the LS. The N terminus of the LS has been reported as Thr-5 and Ala-15 in tobacco (3), Ala-15 in barley (4), Ser-2 in wheat (5), and acetylated Pro-3 in spinach (6). The N termini of the LS of RbuP\(_2\) carboxylase from wheat (5) and two species of 

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: RbuP\(_2\), ribulose 1,5-bisphosphate; LS, large subunit of RbuP\(_2\) carboxylase; SS, small subunit of RbuP\(_2\) carboxylase; FAB, fast-atom bombardment; MS, mass spectrometry; CID, collision-induced dissociation; u, atomic mass unit(s).
ketone-treated), and trypsin inhibitor (type II-0) were purchased from Sigma. Trifluoroacetic acid (sequential grade) and constant-boiling 6 M HCl (sequential grade) were obtained from Pierce Chemical, and HPLC grade acetonitrile was from Fisher. All aqueous solutions were prepared with Milli-Q (Millipore) deionized water (18 MΩ cm−1).

RESULTS

Limited trypic proteolysis of RbuP2 carboxylase from spinach, wheat, tobacco, and muskmelon released several terminal peptides derived from the LS as judged by small molecular weight changes after denaturing NaDodSO4/PAGE on a concave 10–20% gradient gel (data not shown, but see ref. 6). There were no molecular weight changes in the SS after proteolysis, as judged by unaltered electrophoretic mobility of the SS. Fractionation by HPLC of tryptic peptides released from the LS of RbuP2 carboxylase from spinach, wheat, tobacco, and muskmelon revealed several peptides (Fig. 1). Two of these peptides were collected, purified to homogeneity, and characterized (peaks 1 and 2, Fig. 1).

The first peptide eluted by gradient HPLC (peak 1, Fig. 1) of the tryptic digests of the enzymes from wheat, tobacco, and muskmelon had retention time identical to the one from spinach RbuP2 carboxylase. This peptide was identified (6) as the N-terminal tryptic peptide, acetyl-Pro-3 to Lys-8. Each of the peptides derived from RbuP2 carboxylase from wheat, tobacco, and muskmelon contained blocked N termini; no phenylthiohydantoin derivatives were released after five cycles of Edman degradative sequencing. Amino acid composition analyses of these peptides yielded similar results to the peptide derived from spinach RbuP2 carboxylase (Fig. 2). FAB-MS confirmed that these peptides were identical, with parent ion molecular weights of M + H+ [m/z = 745.4 atomic mass units (u)] and matching daughter spectra obtained by CID (for daughter spectra, see figure 5 of ref. 6). Both the parent ion and all N-terminal CID fragments were shifted in mass by 42 u from that of a free amino group, suggesting an acetylated N terminus. Unequivocal demonstration of the N-terminal blocking moiety was established by an exact molecular mass determination of the LS N-terminal tryptic peptide from spinach RbuP2 carboxylase, thus defining the elemental composition of the peptide. The measured mass for this peptide (745.3724 u) and the calculated mass (745.3731 u) based on the expected amino acid sequence with the inclusion of an N-terminal acetyl modification were in close agreement. Thus, we conclude that the LS N terminus of spinach, wheat, tobacco, and muskmelon RbuP2 carboxylase is acetylated Pro-3.

The second tryptic peptide (peak 2, Fig. 1) from the LS was not identical for the four preparations. The peptide from the spinach or wheat enzyme was readily sequenced by Edman degradation and FAB-MS. The results substantiated earlier reports of the amino acid composition of the wheat N terminus (5) and sequencing analysis of the spinach N terminus (6), which identified this peptide as the penultimate N-terminal tryptic peptide of the LS of RbuP2 carboxylase. Position 10 in the LS of RbuP2 carboxylase from spinach, tobacco, and muskmelon leaves was a serine residue, but this was substituted by glycine in the enzyme from wheat. The molecular weight of the spinach penultimate N-terminal LS tryptic peptide was M + H+ [m/z = 608.3 u] and the corresponding wheat peptide was M + H+ [m/z = 578.3 u]. The CID daughter spectra for each peptide is in agreement with earlier analyses (spectra not shown).

In contrast to the enzyme from spinach and wheat the penultimate N-terminal regions of the LS of RbuP2 carboxylase from tobacco and muskmelon, although identical, differed from the peptides obtained from spinach and wheat. Edman degradative sequencing of these peptides demonstrated identical sequences beginning with alanine as the N-terminal residue and ending with the expected C-terminal lysyl residue.

Ala-Ser-Val-Gly-Phe-Xaa-Ala-Gly-Val-Lys

9 14 18

However, during the sixth cycle of Edman degradative sequencing no identifiable phenylthiohydantoin derivatives were obtained. In tobacco the DNA sequence for the LS codes for a lysyl residue at position 14 (14), which has been substantiated in one report by direct sequencing (3). However, in our studies this site was not susceptible to proteolytic cleavage by trypsin.

The parent ion molecular weight for this tryptic peptide from the tobacco or muskmelon enzyme was M+ [m/z = 1005.6 u], which was considerably larger than the penultimate N-terminal tryptic peptide derived from the spinach and
wheat enzymes. In contrast it eluted during HPLC within 0.5 min of the corresponding spinach peptide. The CID daughter spectra of the parent ion for this peptide, although identical from the two plants, was unusual (Fig. 3). The structure of this peptide prevented formation of lower mass fragment ions (<600 u). This suggested that the central region of this peptide contained a unique structural modification. Based on the reported DNA sequence for the LS of tobacco RbuP2 carboxylase (14) and the CID fragmentations, the amino acid at position 14 of this peptide is shifted in mass by 42 u from the expected lysyl residue.

The identity of the unknown amino acid residue located at position 14 in these peptides was suggested by amino acid composition analyses. The amino acid composition of the peptides from tobacco or muskmelon was identical, each contained 1 mol equivalent of trimethyllysine as determined by co-elution with authentic N\textsuperscript{3}-trimethyllysine (Fig. 4). The presence of a trimethyllysyl residue at position 14 would readily explain the absence of tryptic proteolysis, the 42-u molecular mass shift, and the inability to cleave past this position in the CID daughter spectrum due to the quaternary amine. The presence of a N\textsuperscript{3}-trimethyllysyl residue was unequivocally established by an exact molecular mass determination. The measured mass is M+ rather than M + H+ due to the quaternary ammonium group. The calculated mass for a peptide with the primary sequence from Ala-9 to Lys-18 containing a trimethyllysyl residue at position 14 is 1005.6100 u. The measured mass from three separate determinations was 1005.6108 u, 1005.6118 u, and 1005.6088 u. Other modifications that could result in a 42-u mass shift result in substantially different masses, such as a propyl (1005.6022 u) or an acetyl (1005.5730 u) group. Thus, the modified residue at position 14 in the N-terminal region of the LS from tobacco and muskmelon RbuP2 carboxylase is a trimethyllysyl residue.

Proteolytic removal of the penultimate N-terminal region from Ala-9 to Lys-14 in the LS of spinach RbuP2 carboxylase correlates with catalytic inactivation (R.L.H. and R.M.M., unpublished results). Removal of the similar but larger tryptic peptide from Ala-9 to Lys-18 from the LS of muskmelon and tobacco RbuP2 carboxylase should also result in catalytic inactivation. Under identical conditions, limited tryptic proteolysis of spinach or muskmelon RbuP2 carboxylase resulted in irreversible catalytic inactivation, but the rate of inactivation of the muskmelon enzyme was reduced by a factor of 8 compared to the rate for the spinach enzyme (Fig. 5).

Thus, although tryptic proteolysis cannot occur at position 14 in the N terminus of the LS of muskmelon RbuP2 carboxylase, due to trimethyllysine at this position, a slower rate of proteolysis does occur at Lys-18.

The location of tryptic sensitive lysyl residues, post-translational modifications, and the sequence for the N-terminal region of the LS of RbuP2 carboxylase from spinach, wheat, tobacco, and muskmelon is summarized in Fig. 6.
DISCUSSION

Limited tryptic proteolysis of RbuP2 carboxylase isolated from spinach, tobacco, wheat, and muskmelon released two adjacent N-terminal tryptic peptides from the LS. The primary structure of the N-terminal tryptic peptide was identical to all species, as established by: Edman degradative sequencing; amino acid composition analyses; and tandem MS. Exact molecular mass measurement of the spinich tryptic peptide unequivocally established the N termini of the LS from these RbuP2 carboxylases as acetylated Pro-3. Comparisons with the available known DNA sequences for the LS of RbuP2 carboxylase from wheat (8), tobacco (14), and spinach (15) allow the deduction that the N termini have all been post-translationally processed by removal of Met-1 and Ser-2 followed by acetylation of Pro-3. Although the DNA sequence is not available for RbuP2 carboxylase from muskmelon, it is reasonable to expect identical processing for the LS of the muskmelon enzyme.

The penultimate N-terminal tryptic peptide released from the LS of spinach or wheat RbuP2 carboxylases was readily sequenced by tandem MS, which confirmed earlier results obtained with conventional techniques. However, the peptide from the LS of RbuP2 carboxylase from tobacco and muskmelon was post-translationally modified and produced an unexpected mass (\(M^+ = 1005.6\) u). Tandem mass spectrometry of \(M^+ = 1005.6\) u gave a spectrum (Fig. 3) that was consistent with the peptide Ala-9 to Lys-18 plus an additional 42 u. The CID daughter spectrum agreed with this sequence with the extra 42 u occurring at Lys-14. Furthermore, the CID daughter spectrum gave an unusual fragmentation pattern, showing essentially no fragment ions below \(m/z = 600\) u. This was due to the presence of a trimethyllysyl residue bearing a permanent positive charge that prevented the formation of fragment ions that did not contain that charge site. The presence of trimethyllysine also resulted in the absence of any clearly defined phenylthiohydantoin residue released during the sixth cycle of Edman degradative sequencing, although the remainder of the sequence was clear. Ironically, a simple amino acid composition analysis gave the first indication that the amino acid residue at position 14 was trimethyllysine. The FAB-MS exact mass measurements provided conclusive evidence that the 42-u mass shift in the parent ion molecular weight of the penultimate N-terminal peptide from the LS of RbuP2 carboxylase from tobacco and muskmelon was due to trimethylation of the e-amino group of Lys-14.

The penultimate N-terminal region from Ala-9 to Lys-14 in the LS of RbuP2 carboxylase from spinach and wheat is required for normal levels of catalytic activity. Thus, during limited tryptic proteolysis, LS N-terminal peptides are sequentially released and RbuP2 carboxylase is inactivated. An earlier report investigating the effect of trypsin on RbuP2 carboxylases from several plant species found that the tobacco enzyme was resistant to tryptic proteolytic inactivation compared to spinach (5). These authors speculated that given nearly identical primary structures, the resistance of the tobacco enzyme to proteolytic inactivation could be due to differences in tertiary structure. RbuP2 carboxylase from muskmelon is also resistant to tryptic proteolytic inactivation compared to the spinach enzyme. However, the cause of this difference in proteolytic inactivation is probably not due to differences in tertiary structure, but the absence of proteolysis at trimethyl-

---

**Fig. 5.** Tryptic proteolytic inactivation of spinach and muskmelon RbuP2 carboxylase. Activated spinach (●) or muskmelon (●) RbuP2 carboxylase (2 mg/ml) was proteolyzed with trypsin (10 \(\mu g/ml\)) and aliquots were assayed for carboxylase activity (A) at various intervals. Kinetic analysis (B) for irreversible enzyme inactivation.

**Fig. 6.** N terminus of the LS from RbuP2 carboxylase. Solid lines and arrows denote tryptic peptide fragments and proteolytic cleavage sites, respectively, with parent ion molecular weights below each fragment. Boxes denote amino acid residues that vary between species.
lysine-14, and a subsequent slower rate of proteolysis at Lys-18. Thus, trimethyllysine is not recognized by trypsin (18), and the tryptic accessibility of Lys-18 in muskmelon, and by inference Lys-18 in the tobacco enzyme, is less than Lys-14 in the spinach and wheat enzymes.

Trimethyllysine is a relatively rare amino acid, but has been documented in cytochrome c, calmodulin, ribosomal proteins, and histones (19, 20). Calmodulin N-methyltransferase is highly specific for transfer of methyl groups from S-adenosylmethionine to the ε-amine of lysine-115 in calmodulin (21). The N-methyltransferases operative on ribosomal proteins (22) and histones (23) are also specific. Although the RbpP2 carboxylase LS N-methyltransferase has yet to be isolated, it too may have a high degree of specificity, since the substrate lysine-14 residue is flanked by adjacent nonmethylated lysyl residues, one as close as position 18.

Our results provide more detail on the primary structure of the LS from RbpP2 carboxylase isolated from spinach, wheat, tobacco, and muskmelon. However, a functional role for N-terminal acetylation or N2-trimethylation in other proteins is largely unknown. Proteins with acetylated N termini are common but a clear role for this post-translational modification is absent (24). In isolated cases, such as β-endorphin (25) and α-melanocyte-stimulating hormone (26), N-terminal acetylation directly affects activity. A possible protective role against proteolysis has been debated (27). N-terminal proteolytic processing and acetylation in the LS of RbpP2 carboxylase may be related to holoenzyme assembly, perhaps as a prerequisite for association with a chloroplastic chaperonin (28), although other functions are equally possible.

Comparisons between properties of the enzyme from muskmelon and tobacco against the spinach and wheat enzymes cannot be used to elucidate the function of N2-trimethylation, because the primary sequences of these enzymes are not entirely identical. Therefore, properties of the enzyme from tobacco and muskmelon, such as ease of crystallization, lower specific activity, and in vivo regulation by carboxyarabinitol 1-phosphate (29), may be related to but not solely attributed to the presence of trimethyllysine. One functional aspect of trimethylation is suggested from the methodology used to obtain N-terminal peptides from the LS of RbpP2 carboxylase. Trimethyllysine is not recognized by trypsin, and lysine-specific proteases may exist in plants. In this respect, trimethylation would provide protective protection to the LS of RbpP2 carboxylase, although other nonmethylated N-terminal lysyl residues are present. However, trimethylation of lysine-115 in calmodulin does provide protection against the specific ATP/ubiquitin-dependent proteolytic system (30).

X-ray crystallography of tobacco RbpP2 carboxylase revealed that the electron density of the N-terminal region of the LS from the N termini to residue 19 did not map well (10). Thus, the role of this region in catalysis and the consequential effects of trimethylation remain obscure. Since the N-terminal region of the LS of RbpP2 carboxylase in some species is particularly susceptible to proteolysis during isolation (31) and removal of this region severely reduces catalytic activity (5, 8), structural studies should be used routinely to verify the intactness of the LS N-terminal region. Proteolyzed forms of RbpP2 carboxylase whose LS penultimate N-terminal regions have been removed are catalytically deficient but capable of activator CO2 carbamate formation (5) and reaction-intermediate analogue binding (6).

We thank Malcolm Royer for technical assistance and Bill Henzel and Steve Gathy for gas-phase sequencing data. Initial mass spectral results were obtained at the NIH/MSU Mass Spectrometry Facility, supported by a grant from the Research Resources Division of the National Institutes of Health (RR-00480). This research was supported by the Kentucky Agricultural Experiment Station, Hatch Project KY 00586 to R.L.H. and is published as Kentucky Agricultural Experiment Station Article 88-10-269.