Thiaminase I (42 kDa) Heterogeneity, Sequence Refinement, and Active Site Location from High-Resolution Tandem Mass Spectrometry

Neil L. Kelleher, Colleen A. Costello, Tadhg P. Begley, and Fred W. McLafferty

Department of Chemistry, Baker Laboratory, Cornell University, Ithaca, New York, USA

Thiaminase I (E.C. 2.5.1.2) from *Bacillus thiaminolyticus* catalyzes the degradation of thiamin (vitamin B_1). Unexpected mass heterogeneity (MW 42,127, 42,197, and 42,254; 1:2:1) in recombinant thiaminase I from *Escherichia coli* was detected by electrospray ionization Fourier-transform mass spectrometry, resolving power 7×10^4. Nozzle–skimmer fragmentation data reveal an extra Ala (+71.02; 71.04 = theory) and GlyAla (+128.04; 128.06 = theory) on the N-terminus, in addition to the fully processed enzyme. However, the fragment ion masses were consistent only with this sequence through 330 N-terminal residues; resequencing of the last 150 bps of the thiaminase I gene yields a sequence consistent with the molecular weight values and all 61 fragment ion masses. Covalently labeling the active site with a 108-Da pyrimidine moiety via mechanism-based inhibition produces a corresponding molecular weight increase in all three thiaminase I components, which indicates that they are all enzymatically active. Inspection of the fragment ions that do and do not increase by 108 Da indicates that the active site nucleophile is located between Pro^{79} and Thr^{177} in the 379 amino acid enzyme. *(J Am Soc Mass Spectrom 1995, 6, 981–984)*

For proteins as large as 67-kDa [1], electrospray ionization (ESI) [2] combined with Fourier-transform mass spectrometry (FTMS) [3] provides unusual resolving power and tandem mass spectrometry (MS^n) capabilities that yield accurate molecular weight and partial sequence information [1, 4, 5]; previously such data have been obtained only for well-characterized proteins. Thiaminase I (E.C. 2.5.1.2) from *Bacillus thiaminolyticus* catalyzes the degradation of thiamin (vitamin B_1) by displacement of the thiazole moiety with a variety of nucleophiles (Scheme I) [6], and thus excess thiaminase I can cause beriberi [7]. Initial characterization of recombinant thiaminase I from *Escherichia coli* by DNA sequencing, sodium dodecylsulfate–polyacrylamide gel electrophoresis (SDS-PAGE), and amino-terminal sequencing indicated high purity (Costello, C. A.; Kelleher, N. L.; Abe, M.; McLafferty, F. W.; Begley, T. P., unpublished; hereafter CKAMB).

Experimental

The thiaminase I gene was cloned into the *NdeI-SalI* sites of the pET22b(+) expression vector, which was used to transform *E. Coli* strain BL21(DE3), as described separately (CKAMB). Thiaminase I was isolated from cell lysate by ammonium acetate precipitation, followed by anion exchange and dye column chromatography (CKAMB). DNA and Edman sequencing were performed by the Cornell DNA sequencing facility. Purified thiaminase I in phosphate storage buffer (15% glycerol) was desalted via ultrafiltration (Amicon 10), diluted to ~10 μM in 76:20:4 $MeOH/H_2O/HOAc$, and electrosprayed at 1 μL/min through a needle (2.8 kV) and capillary (~110 °C). The resultant ions pass through three rf-only quadrupoles into a cylindrical trapped ion cell (10^{-9} torr) of a 6-T Millipore-Extrel (Pittsburgh, PA) FTMS 2000, described elsewhere [8]. Mass values reported are those for the most abundant ^{13}C peak (designated in parentheses) in the isotopic distribution.

Results and Discussion

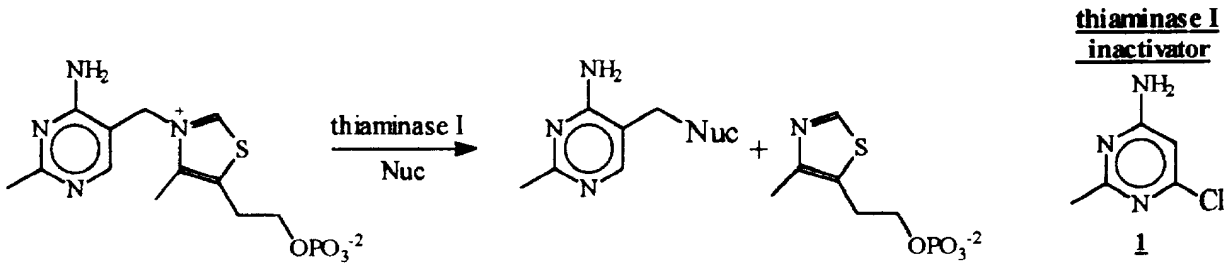
Heterogeneity

Electrospray ionization/Fourier-transform (ESI/FT) mass spectra of recombinant thiaminase I unexpectedly reveal three species of MW = 42,127, 42,197, and 42,254 (all $^{13}C_{26}$) in a 1:2:1 ratio with 7×10^4 resolving power (RP; Figure 1). A seriously misleading molecular weight value would result without resolution of this heterogeneity; this requires RP ~600 without salt

Address reprint requests to Fred W. McLafferty, Baker Laboratory, Department of Chemistry, Cornell University, Ithaca, NY 14853-1301.

1044-0305/95/$9.50
SSDI 1044-0305(95)00584-Z

Revised August 1, 1995
Accepted August 2, 1995



Scheme I.

adducts, well above that of SDS-PAGE, and not yet demonstrated for matrix-assisted laser desorption time-of-flight or ESI-quadrupole mass spectrometry in this mass range [9]. These molecular weight values (accuracy ±1 Da) indicate that thiaminase I contains approximately 30 more amino acids than predicted by the DNA sequence determined by extensive bidirectional Sanger sequencing (CKAMB). Although the N-terminal signal peptide (~28 aa) could account for this difference exactly, its removal was shown by Edman sequencing that indicated N-terminal AAH residues.

Nozzle–skimmer (NS) fragmentation of the heterogeneous molecular ions (Figure 2) gave resolved isotopic patterns, which enabled direct charge state (z) determination and mass (m) assignment [10] for 165 fragment ions of 120 mass values. Many fragment ions had only one charge state and thus would be uninterpretable on lower resolution instruments. Two sets of complementary ion pairs [5] were found that sum to 42,127 Da: 33,355.4 ($^{13}C_{21}$) + 8771.87 ($^{13}C_5$) and 36,146.0 ($^{13}C_{23}$) + 5981.17 ($^{13}C_3$); the components of each ion pair should contain the N- and C-termini [5], denoted as b and y ions, respectively [11]. The 8771.87 and 5981.17 ions (but not their partner ions) both have satellites at +71.02 Da and +128.04 (Figure 2b) that correspond to the three molecular species, which indicates that they have extensive sequence homology. The

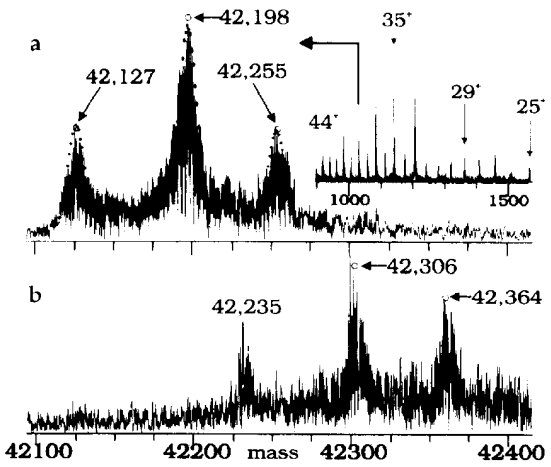


Figure 1. (a) 39+ molecular ion region of a (inset) broadband ESI/FT mass spectrum of recombinant thiaminase I, 20 scans; dots, theoretical isotopic distributions. (b) 38+ region after 24-h incubation of thiaminase I with 4-amino-6-chloro-2-methylpyrimidine.

NS spectrum shows 12 other fragment ions with satellites at +71.0 ± 0.1 and +128.0 ± 0.1 Da, which include a string of four from 8381.66 to 8771.87 that yields the partial sequence Y-L(I)-L(I) (Figure 2a). The DNA-derived sequence contains the corresponding string $Y^{76}L^{77}L^{78}$ that is consistent with the fragment ion masses if the Edman-indicated N-terminus is assigned instead as $A^{-1}A^{1}H^{2}$; the satellite ions then arise from the extra A^{-1} (71.04 Da) and extra $G^{-2}A^{-1}$ (128.06) on the N-terminus (Figure 3). The original Edman sequencing stopped after three cycles and did not give clear evidence of the GAA and AHS from the minor components in the presence of AAH (note the amino acid degeneracy and that G is often a contaminant in the first sequencing cycle). All 14 fragment ions that exhibit heterogeneity (the b ions) are consistent with this NH_2-terminal assignment through residue 78. The N-terminal heterogeneity is attributed to a semispecific enzymatic removal of the signal peptide (CKAMB and [12]).

Sequence Refinement

Of the remaining non-b (no satellites) ions, the 33,355.4 and 36,146.0 Da must be y ions, because they are complements of b_{78} and b_{53}; thus their 2790.6 ($^{13}C_2$) difference corresponds accurately to that expected (2790.4) for the DNA-predicted sequence P^{54}-L^{78}. Of the remaining fragment ions, 14 others had mass values whose differences were similarly consistent with the DNA-predicted sequence through Y^{330} (Figure 3), with seven additional ions that correspond to internal fragments. Subsequent resequencing of the 3'-region of the thiaminase I gene (~150 bases) found an extra base at the codon for residue 334, which yields a sequence that predicts three species of 379 (42,126), 380 (42,197), and 381 residues (42,254) that are within ±1 Da of the observed molecular weight values and are consistent with the mass values of 61 fragment ions (CKAMB).

Active Site Location

The mass spectrum of thiaminase I inactivated with a known mechanism-based inhibitor, 4-amino-6-chloro-2-methylpyrimidine (**1**, Scheme I) [13], shows a 108 ±

1 Da shift for each component, evidence that all components are enzymatically active and have only one pyrimidine binding site. NS fragmentation shows 42 fragment ions that are not shifted 108 Da by the inhibitor and two fragments, the y_{301} and the internal $b_{177}y_{301}$, that are shifted (Figure 3) 107.8 and 108.3, respectively (other fragments were of insufficient abundance). Thus the 99 residue region between Pro^{79} and Thr^{177} should contain the active site nucleophile, providing the first evidence of its correct location.

Reflecting a parallel study (Wood, T. D.; Kelleher, N. L.; Little, D. P.; Chen, L. H.; Kenyon, G. L.; McLafferty, F. W., unpublished), here ESI/FTMS has obtained unique and detailed information on the purity, sequence, and active site of an intact large protein that is fully confirmed by current biological studies to be reported in the full paper (CKAMB). Demonstrating a further unique advantage, mass spectra of the quality similar to Figures 1 and 2 can now be obtained on subfemtomole samples [14].

Acknowledgments

The authors thank Troy D. Wood, Gary Valaskovic, Daniel P. Little, Michael W. Senko, Peter B. O'Connor, and the Cornell DNA sequencing facility for experimental assistance and helpful ideas, and the National Institutes of Health for generous funding (NLK, Cell & Molecular Biology training grant 08-T2GM07273; FWM, GM16609; TPB, DK44083).

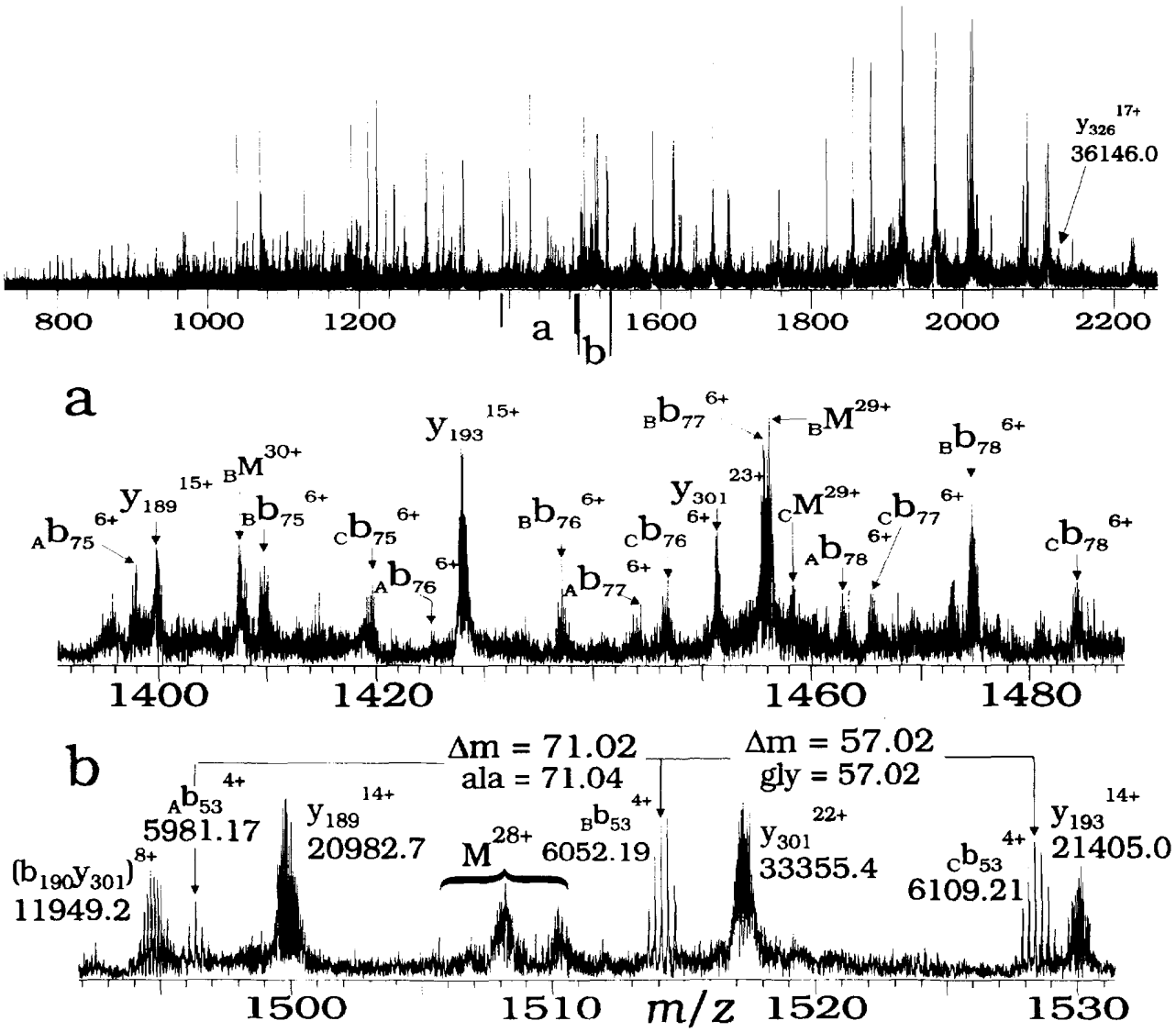


Figure 2. Nozzle–skimmer (130 V) dissociation spectrum of thiaminase I, 30 scans; preceding subscripts A, B, or C denote origin from the smallest, middle, or largest molecule species, respectively.

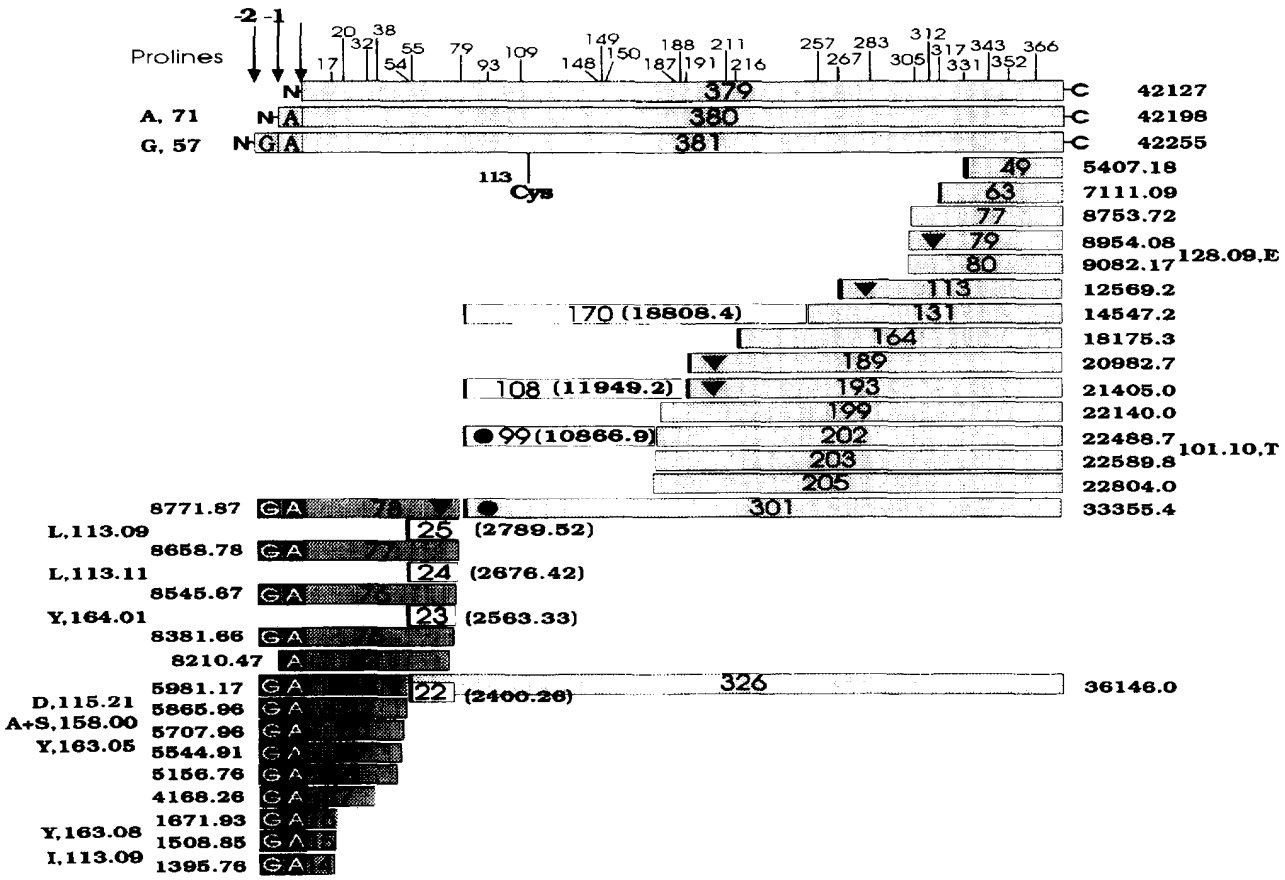


Figure 3. Thiaminase I fragmentation data; y ions on right; b ions on left. Upper left arrows, post-translational cleavage sites. ●, fragment shifts + 108 Da upon inactivation; ▼ fragment does not shift.

References

1. Speir, J. P.; Senko, M. W.; Little, D. P.; Loo, J. A.; McLafferty, F. W. *J. Mass Spectrom.* **1995,** *30,* 39–42.
2. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Mass Spectrom. Rev.* **1990,** *9,* 37–70.
3. Marshall, A. G.; Grosshans, P. B. *Anal. Chem.* **1991,** *63,* A215–A229.
4. McLafferty, F. W. *Acc. Chem. Res.* **1994,** *27,* 379–386.
5. O'Connor, P. B.; Speir, J. P.; Senko, M. W.; Little, D. P.; McLafferty, F. W. *J. Mass Spectrom.* **1995,** *30,* 88–93.
6. Fujita, A. *Adv. Enzymol. Related Subjects Biochem.* **1954,** *15,* 389–421; Murata, K. *Ann. N.Y. Acad. Sci.* **1982,** *378,* 146–156.
7. Earl, J. W.; McCleary, B. V. *Nature* **1994,** *368,* 683–684.
8. Beu, S. C.; Senko, M. W.; Quinn, J. P.; Wampler, F. M., III; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1993,** *4,* 557–565.
9. Biemann, K.; Papayannopoulos, I. A. *Acc. Chem. Res.* **1994,** *27,* 370–378.
10. Senko, M. W.; Beu, S. C.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1995,** *6,* 52–56.
11. Roepstorff, P.; Fohlman, J. *Biomed. Mass Spectrom.* **1984,** *11,* 601.
12. Hitzeman, R. A.; Leung, R. A.; Perry, L. J.; Kohr, W. J.; Levine, H. L.; Goeddel, D. V. *Science* **1983,** *219,* 620–625; Lingappa, V. R.; Devillers-Thiery, A.; Blobel, G. *Proc. Nat. Acad. Sci.* **1977,** *74,* 2432–2436.
13. Hutter, J. A.; Slama, J. T. *Biochemistry* **1987,** *26,* 1969–1973.
14. Valaskovic, G. A.; Kelleher, N. L.; Little, D. P.; Aaserud, D. J.; McLafferty, F. W. *Anal. Chem.*, in press.