Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 1947–1949, March 1990
Biochemistry

Sequencing of peptides and proteins with blocked N-terminal amino acids: *N*-Acetylserine or *N*-acetylthreonine

(Edman degradation/amino acid sequence/deblocking/N-acetylated amino acids/phenyl isothiocyanate)

DANIEL WELLNER, C. PANNEERSELVAM*, AND B. L. HORECKER

Department of Biochemistry, Cornell University Medical College, New York, NY 10021

Contributed by B. L. Horecker, December 29, 1989

ABSTRACT **Many proteins cannot be directly sequenced by Edman degradation because they have a blocked N-terminal residue. A method is presented for deblocking such proteins when the N-terminal residue is *N*-acetylserine (which occurs frequently in eukaryotic proteins) or *N*-acetylthreonine. The method has been applied successfully to the determination of the N-terminal amino acid sequence of human, bovine, and rat parathymosins. Prothymosin α and other blocked proteins and peptides were also readily deblocked and sequenced by this procedure. It is proposed that the mechanism of the deblocking reaction involves an acid-catalyzed N → O shift of the acetyl group followed by a β-elimination.**

Although notable improvements have been made recently in the instrumentation available for automated sequencing of proteins and peptides, many proteins still present a challenging problem to investigators who attempt to determine their sequence. One problem frequently encountered is that the N-terminal residue is modified in such a way that it does not react with the Edman reagent phenyl isothiocyanate. For example, the blocked N-terminal residue may be an *N*-acetyl amino acid (1), a glycosylated amino acid (2), or a pyrrolidone carboxylate group (3). Of these, proteins with an N-acetylated amino acid are encountered most frequently. Evidence has been presented that about 80% of the soluble proteins in mammalian cells have acetylated N-terminal amino acids (4).

Previous attempts to remove the blocking group have used enzymatic or limited acid hydrolysis. Nakamura *et al.* (5) have used a rat liver peptidase to remove *N*-acetylserine from the N-terminal peptide released from thrombin by bovine Factor XIII. Both the rat liver peptidase (6) and a similar enzyme from human erythrocytes (7) were shown to split off *N*-acetylserine from α-melanocyte-stimulating hormone. However, the applicability of this method is limited by the restricted specificity of these enzymes. In particular, they do not appear to act efficiently on large peptides or proteins. Proteins blocked with a pyrrolidone carboxylate group often can be unblocked by treatment with pyrrolidone carboxylate peptidase (8). Some success has been reported in removing *N*-acetyl groups from peptides by limited HCl hydrolysis (9, 10).

In this report, a nonenzymatic method is presented for deblocking proteins and peptides having N-terminal acetylserine or acetylthreonine residues. This method has been applied successfully to the elucidation of the primary structure of prothymosin α and parathymosin, acidic polypeptides found in the thymus gland and other mammalian tissues.

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.

MATERIALS AND METHODS

Materials. Human, bovine, and rat parathymosins were isolated from liver as described (11). Bovine and rat prothymosins α were isolated from thymus glands (11). Thymosin β_4 was purified from rat thymus (12). Horse heart cytochrome *c* was from Schwarz/Mann. *Drosophila* aldolase (13) was a gift from O. Brenner-Holzach, and the synthetic N-acetylated peptides were gifts from E. P. Heimer of Hoffmann–La Roche. The reagents and supplies used for sequencing were purchased from Applied Biosystems.

Deblocking Proteins for Sequencing. A trifluoroacetic acid-treated glass fiber filter disc, 12 mm in diameter, was inserted slightly folded into a 1.5-ml polypropylene microcentrifuge tube. The filter was wetted with 30 μl of a solution containing Polybrene (3 mg) and NaCl (0.2 mg) and dried. A solution of the protein to be tested was applied to the filter and dried. The filter was then saturated with 30 μl of anhydrous trifluoroacetic acid (sequencer reagent 3), and the tube was closed. After incubating for 4 min at 45°C, the tube was opened in the hood to allow most of the trifluoroacetic acid to evaporate. After 5 min at room temperature, the open tube was allowed to dry for another 10 min at 45°C. The tube was then closed and placed in an oven at 65°C for 16 hr or at 45°C for 3 days. Sequencing was then carried out according to the manufacturer's recommendations in an Applied Biosystems gas-phase sequencer model 470A equipped with an on-line phenylthiohydantoin (PTH) analyzer model 120A. Program O3RPTH was used.

RESULTS

It was found that several proteins that contain an N-acetylated serine residue, after treatment with anhydrous trifluoroacetic acid, undergo a time-dependent reaction in which they become deblocked and amenable to sequencing by the Edman method. Conditions that have been found to work well for a number of proteins are described under *Materials and Methods*. The protein is first treated with anhydrous trifluoroacetic acid for 4 min at 45°C. After removal of the acid, the treated protein is incubated at 65°C for 16 hr or for 3 days at 45°C. Initial yields vary, depending on the protein. Yields ranging from 3% to 40% have been obtained.

Table 1 shows the N-terminal sequence of rat parathymosin obtained by this method. The initial yield was about 7%. The N-terminal serine had previously been shown to be acetylated (11). An identical N-terminal sequence was obtained for bovine parathymosin by using the same method for deblocking the protein (14). The "background" in this se-

Abbreviation: PTH, phenylthiohydantoin.

*Present address: Department of Biochemistry, Post Graduate Institute of Basic Medical Sciences, Madras University, Madras 600113, India.

Table 1. N-terminal sequence of rat parathymosin

Position	1	2	3	4	5	6	7
Residue identified	Ser	Glu	Lys	Ser	Val	Glu	Ala
Amount, pmol	97*	115	55	87*	47	54	50
Yield, %	6.5	7.7	3.7	5.8	3.1	3.6	3.3

Parathymosin (1.5 nmol) was deblocked for 3 days at 45°C as described. The N-terminal residue was shown previously to be acetylated (11).
*Calculated from the sum of PTH-serine and the dithiothreitol adduct of PTH-dehydroalanine.

quencing run was low, and it was possible to sequence the first 20 residues. The same N-terminal sequence was also obtained for human parathymosin and subsequently confirmed by sequencing the cloned cDNA coding for this protein (15).

A number of other blocked proteins and peptides have been successfully sequenced by this method. These include bovine and rat prothymosin α, rat thymosin β_4, and synthetic *N*-acetylserine- and *N*-acetylthreonine-containing peptides.

The application of this procedure to *N*-acetylthreonine-containing peptides is shown in Table 2. An 11-residue peptide was sequenced in its entirety with an initial yield of about 6%. A separate sequencing run carried out without deblocking showed that at least 99.5% of the peptide was blocked.

In experiments with thymosin β_4, when the reaction time at 65°C was varied, it was found that the yield was optimal at about 16 hr (Table 3). Under these conditions, the initial yield was about 40%, substantially higher than with other proteins tested. The yields of two representative amino acids, lysine at position 3 and proline at position 4, are shown in Table 3. There was no increase in yield between 16 and 24 hr.

Examination of the background peaks in the sequencing of thymosin β_4 and other proteins after deblocking suggested that some cleavage occurred at serine and threonine residues in the interior of the peptide chain. However, the extent of cleavage of these internal peptide bonds is considerably less than that of the N-terminal acetyl group. An estimate of the relative amount of internal and N-terminal cleavage may be obtained by measuring the yield of background PTH-glutamate at cycle 3 of the sequence. Thymosin β_4 is a 43-residue peptide containing three serine and three threonine residues in addition to the N-terminal acetylserine (16). The PTH-glutamate peak is relatively prominent in the background at cycle 3 because three of these hydroxyamino acid residues (one serine and two threonines) occur two positions before a glutamate residue. If, during the deblocking procedure, the peptide bond involving the amino group of one of these serine or threonine residues is cleaved, PTH-glutamate will appear at cycle 3. As shown in Table 3, the average background PTH-glutamate peak at position 3 corresponds to about 30% of the average PTH-lysine peak at position 3. If one assumes equal cleavage at all internal serine and threonine residues during deblocking, this would indicate that each internal cleavage corresponds to about 10% of the cleavage of the N-terminal acetyl group (Table 3, bottom line). Therefore, it may be concluded that the N-terminal acetyl linkage is considerably more labile than any internal peptide bond under the conditions used.

Table 3. Sequencing yields with thymosin β_4 after different reaction times at 65°C

Incubation time at 65°C, hr	1	4	10	16	24
Lys-3 yield, %	4.6	10.0	10.6	41	41
Pro-4 yield, %	5.1	10.1	10.0	37	40
Background Glu-3 yield, %	1.8	2.6	2.5	18	9.2
⅓ Glu-3 yield/Lys-3 yield*	0.13	0.09	0.08	0.15	0.07

Thymosin β_4 (2 nmol) was treated with anhydrous trifluoroacetic acid for 4 min at 45° as described. After drying, samples were placed in an oven at 65°C for the time indicated and sequenced.
*See text.

When horse heart cytochrome *c*, a protein with an *N*-acetylglycine N-terminal, was submitted to the deblocking procedure, no sequence was seen. A low background of PTH–amino acid derivatives was observed, suggesting that some cleavage had occurred at internal peptide bonds. This protein contains 10 threonine and no serine residues (17). Evidence for cleavage on the amino side of 9 of the 10 threonine residues was seen, but the yields were only 0.2–0.6%.

Attempts were also made to sequence *Drosophila* aldolase, a 360-residue polypeptide with an *N*-acetylthreonine terminal group (13). Although the data showed that unblocking had taken place and the protein was being sequenced from the N-terminal end, the background was too high for the sequence to be clearly read. This difficulty may be encountered when applying the method to large proteins, particularly those that are rich in serine and threonine. In such cases, it may be advantageous to isolate an N-terminal peptide prior to deblocking and sequencing.

DISCUSSION

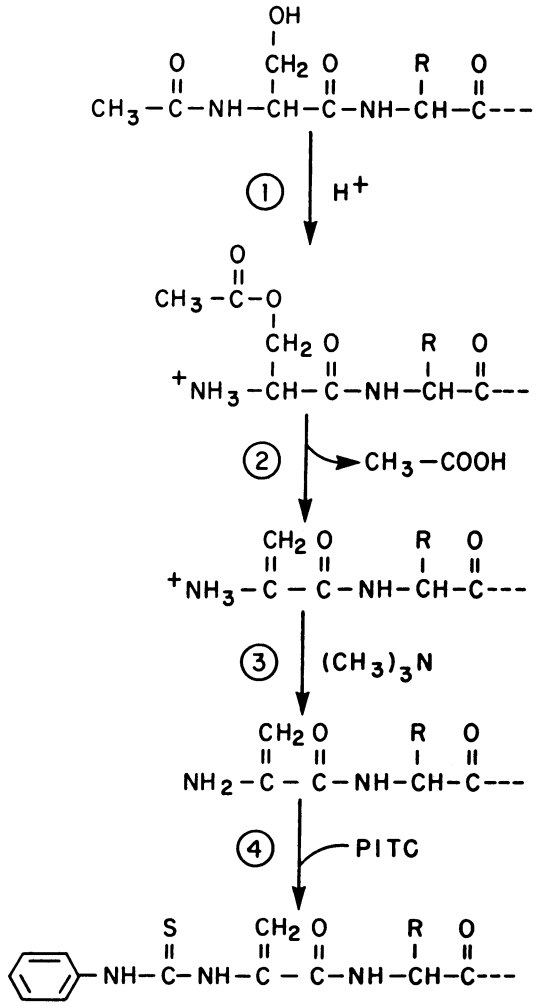
The method described here has been shown to be useful for deblocking proteins and peptides with an N-terminal acetylated serine or threonine. The data show that, after deblocking, the N-terminal sequence of several proteins and peptides could be successfully determined by Edman degradation.

Although the mechanism of the deblocking reaction is not known at this time, the finding that it occurs with proteins having *N*-acetylserine or *N*-acetylthreonine terminal residues but not with cytochrome *c*, which has an *N*-acetylglycine terminus, suggests that an N → O acyl shift may be involved (see Scheme I, reaction 1). Such acyl shifts occur readily under acid conditions and are rapidly reversed when the pH is increased above 7 (18). In addition to an N → O acyl shift, a second reaction must occur before sequencing is started. In the sequencing procedure, the protein or peptide is treated with trimethylamine vapors before and during the reaction with phenyl isothiocyanate. Under these conditions of high pH, the N → O acyl shift would be rapidly reversed, and the amino group would again be blocked. It is possible that the second reaction, which appears to be slower than the N → O transition, is a β-elimination reaction (Scheme I, reaction 2), yielding a protonated dehydroalanyl peptide. In the presence of trimethylamine and phenyl isothiocyanate (PITC in Scheme I), the corresponding phenylthiocarbamoyl derivative may then be formed (Scheme I, reactions 3 and 4). This proposed mechanism is supported by the finding that

Table 2. Sequence of *N*-acetylthreonine synthetic peptide

Position	1	2	3	4	5	6	7	8	9	10	11
Residue identified	Thr	Cys	Asp	Leu	Ala	Pro	Pro	Ala	Gly	Thr	Thr
Amount, nmol	0.79	*	0.74	1.07	1.35	0.90	1.02	1.10	0.66	0.41	0.31
Yield, %	4.0	—	3.7	5.4	6.7	4.5	5.1	5.5	3.3	2.1	1.6

A synthetic peptide (20 nmol) with the sequence *N*-acetyl-Thr-Cys-Asp-Leu-Ala-Pro-Pro-Ala-Gly-Thr-Thr was deblocked and sequenced as described. The incubation was at 65°C for 16 hr.
*The dithiothreitol adduct of PTH-dehydroalanine was seen at this cycle.



Scheme I

most of the PTH derivative observed at the first cycle is that of the dithiothreitol adduct of dehydroalanine. This is in contrast to the finding of a predominant PTH-serine peak usually seen, especially when serine occurs in the early part of a sequence. Although a PTH-serine peak is also seen in the first cycle of deblocked peptides, it may arise by hydration of PTH-dehydroalanine during the conversion step (when the anilinothiazolinone derivative is heated in aqueous acid to form the PTH derivative). Further investigation of this mechanism is needed.

The fact that so many proteins of interest to biochemists and molecular biologists are blocked at the N terminus is hampering efforts to elucidate their structures. Although many different approaches have been used to solve this problem (19), no generally applicable method for unblocking proteins has yet been devised. The present method is restricted to proteins where either serine or threonine is the acetylated N-terminal residue. Nevertheless, it may be applicable to a large number of proteins. A survey of N-acetylated proteins by Persson *et al.* (20) found that, of the known proteins in this class, about 41% have an N-terminal acetylated serine and another 2% an N-terminal acetylated threonine.

This investigation was supported by a Biomedical Research Support Grant and by Grant 1 S10 RR02855-01, both from the National Institutes of Health. Dr. Panneerselvam was the recipient of a Biotechnology Career Fellowship from the Rockefeller Foundation.

1. Narita, K. (1958) *Biochim. Biophys. Acta* **28,** 184–191.
2. Takahashi, N., Takahashi, Y. & Putnam, F. W. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 73–77.
3. Press, E. M., Piggot, P. J. & Porter, R. R. (1966) *Biochem. J.* **99,** 356–366.
4. Brown, J. L. & Roberts, W. K. (1976) *J. Biol. Chem.* **251,** 1009–1014.
5. Nakamura, S., Iwanaga, S., Suzuki, T., Mikuni, Y. & Konishi, K. (1974) *Biochem. Biophys. Res. Commun.* **58,** 250–256.
6. Kobayashi, K. & Smith, J. A. (1987) *J. Biol. Chem.* **262,** 11435–11445.
7. Jones, W. M., Manning, L. R. & Manning, J. M. (1986) *Biochem. Biophys. Res. Commun.* **139,** 244–250.
8. Doolittle, R. F. & Armentrout, R. W. (1968) *Biochemistry* **7,** 516–521.
9. Fordyce, A. M., Midwinter, G. G. & Moore, C. H. (1979) *Biochem. Soc. Trans.* **7,** 721–723.
10. Chin, C. C. Q. & Wold, F. (1985) *Biosci. Rep.* **5,** 847–854.
11. Komiyama, T., Pan, L.-X., Haritos, A. A., Wideman, J. W., Pan, Y.-C. E., Chang, M., Rogers, I. & Horecker, B. L. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 1242–1245.
12. Haritos, A. A., Caldarella, J. & Horecker, B. L. (1985) *Anal. Biochem.* **144,** 436–440.
13. Malek, A. A., Hy, M., Honegger, A., Rose, K. & Brenner-Holzach, O. (1988) *Arch. Biochem. Biophys.* **266,** 10–31.
14. Panneerselvam, C., Clinton, M., Wellner, D. & Horecker, B. L. (1988) *Biochem. Biophys. Res. Commun.* **155,** 539–545.
15. Clinton, M., Frangou-Lazaridis, M., Panneerselvam, C. & Horecker, B. L. (1989) *Biochem. Biophys. Res. Commun.* **158,** 855–862.
16. Low, T. L. K., Hu, S.-K. & Goldstein, A. L. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 1162–1166.
17. Margoliash, E., Smith, E. L., Kreil, G. & Tuppy, H. (1961) *Nature (London)* **192,** 1125–1127.
18. Iwai, K. & Ando, T. (1967) *Methods Enzymol.* **11,** 263–282.
19. Chin, C. C. Q. & Wold, F. (1987) in *Methods in Protein Sequence Analysis–1986*, ed. Walsh, K. A. (Humana, Clifton, NJ), pp. 505–512.
20. Persson, B., Flinta, C., von Heijne, G. & Jörnvall, H. (1985) *Eur. J. Biochem.* **152,** 523–527.