Vol. 157, No. 1, 1988 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
November 30, 1988 Pages 24-29

AMINO ACID SEQUENCE OF THE SIGNAL PEPTIDE OF MITOCHONDRIAL NICOTINAMIDE NUCLEOTIDE TRANSHYDROGENASE AS DETERMINED FROM THE SEQUENCE OF ITS MESSENGER RNA[1,2]

Mutsuo Yamaguchi and Youssef Hatefi[3]
Division of Biochemistry, Department of Basic and Clinical Research
Scripps Clinic and Research Foundation
La Jolla, California 92037

Kathleen Trach and James A. Hoch
Division of Cellular Biology, Department of Basic and Clinical Research
Scripps Clinic and Research Foundation
La Jolla, California 92037

Received October 3, 1988

The amino acid sequence of the bovine mitochondrial nicotinamide nucleotide transhydrogenase was recently deduced from isolated cDNAs and reported [Yamaguchi, M., Hatefi, Y., Trach, K., and Hoch, J.A. (1988) *J. Biol. Chem. 263*, 2761-2767]. The cDNAs lacked the N-terminal coding region, however, and the 8 N-terminal residues were determined by protein sequencing. In the present study, the nucleotide sequence of the 5' upstream region was determined by dideoxynucleotide sequencing of the transhydrogenase messenger RNA, and amino acid sequences of the N-terminal region and the signal peptide of the enzyme were deduced from the nucleotide sequence. The N-terminal sequence of the enzyme as deduced from the mRNA sequence is the same as that determined by protein sequencing, with one difference. Protein sequencing showed Ser as the N-terminal residue. The mRNA sequence indicated that Ser is the second N-terminal residue, and the first is Cys. That preparations of the enzyme are mixtures of two polypeptides, one polypeptide being one residue shorter at the N terminus than the other, has been pointed out in the above reference. The signal peptide consists of 43 residues, is rich in basic (4 Lys, 2 Arg) and hydroxylated (4 Thr, 3 Ser) amino acids, and lacks acidic residues.

The mitochondrial nicotinamide nucleotide transhydrogenase (TH) is an energy-transducing enzyme (1,2), and occurs in the mitochondrial inner membrane as a homodimer (3) of monomer M_r = 110,000 (4). The enzyme catalyzes transhydrogenation between NAD(H) and NADP(H) in a reaction that is coupled to proton translocation across the mitochondrial inner membane (5,6). Recently, the NAD and the NADP binding sites of bovine TH were identified in this laboratory by covalent labeling with tritiated *p*-fluorosulfonylbenzoyl-5'-adenosine (7), and the amino acid sequence of the enzyme was deduced from the cDNA nucleotide sequence (4). However, despite

[1]Supported by United States Public Health Service grants GM24887 and GM19416
[2]This is publication number 5508-BCR from the Research Institute of Scripps Clinic, La Jolla, California.
[3]To whom inquiries should be addressed.

0006-291X/88 $1.50

extensive screening of cDNA libraries, cDNAs encoding the N-terminal region of TH could not be isolated. Hence, the 8 N-terminal residues of the mature protein were determined by protein sequencing. It was found, however, that preparations of TH consisted of a mixture of two polypeptides, one being one residue shorter than the other at the N terminus (4). This problem left the precise identity of the N-terminal residue to be determined. Furthermore, it was known that the TH precursor has a mitochondrial targeting signal peptide with an estimated molecular weight of 2,000 (8) or 5,000 (9), and it was of interest to determine the amino acid sequence of this signal peptide as well.

Recently, a dideoxynucleotide sequencing method has been developed for RNA (10), thus allowing determination of the nucleotide sequence of messenger RNAs. In the present study, the nucleotide sequence of the 5' upstream region of TH mRNA was determined by the above procedure. The results confirmed the N-terminal sequence of the enzyme as was determined by protein sequencing, indicated that the N-terminal residue of the mature TH is Cys, and permitted the amino acid sequence of the signal peptide of TH to be deduced from the mRNA nucleotide sequence.

MATERIALS AND METHODS

Materials - Guanidine thiocyanate was obtained from Fluka; AMV reverse transcriptase was from Stratagene Cloning Systems, San Diego; the Klenow fragment of DNA polymerase I was from Bethesda Research Laboratories; oligo(dT)-cellulose, deoxynucleotide triphosphates, and dideoxynucleotide triphosphates were from Pharmacia; and [γ-^{32}P]ATP (7000 Ci/mmol) was from ICN Radiochemicals.

Preparation of messenger RNA - Bovine heart muscle was diced and homogenized in 4 M guanidinium thiocyanate in a Polytron (Brinkman Instruments, Westbury, NY), and RNA was sedimented through a shelf of 5.7 M cesium chloride (11). Poly (A)$^+$RNA was isolated by oligo (dT)-cellulose chromatography (12).

Synthesis and ^{32}P-labeling of oligonucleotides - Oligonucleotides were synthesized by the Applied Biosystems 380-A DNA synthesizer and purified by polyacrylamide gel electrophoresis under denaturing conditions. Oligonucleotides were ^{32}P-labeled with [γ-^{32}P]ATP and polynucleotide kinase, and the mixture was used without purification for RNA sequencing.

RNA sequencing - The protocol for RNA sequencing developed by Geliebter *et al.* (10) was followed, except that the amount of oligonucleotide primer was reduced from 5 ng to 0.15 ng. Annealing of ^{32}P-labeled oligonuleotide primer and mRNA [Poly (A)$^+$RNA] was carried out 5° C below the calculated denaturation temperature (Td). Td was estimated by the following equation (13): Td = 4(G-C) + 2(A-T). Reverse transcription was carried out at 50° C. For sequencing, 6% polyacrylamide gels containing 7 M urea were used, and the gels were exposed to Kodak XAR-5 films for 4-7 days at -70°C using an intensifying screen. The first oligonucleotide primer (#1, 20 mer) used for sequencing was complementary to the cDNA sequence already determined and, based upon the newly determined sequence, two more oligonucleotides (#2 and 3) were synthesized and further upstream sequences were determined. Using the same primer, sequencing was repeated three times to confirm the sequence data.

Manual Edman degradation - Manual Edman degradation was carried out using dimethylaminoazobenzene isothiocyanate. The dimethylaminoazobenzene thiohydantoin derivatives were identified by two dimensional thin layer chromatography (14).

RESULTS AND DISCUSSION

Nucleotide sequence of the 5' upstream region of the transhydrogenase mRNA - The nucleotide sequence of the 5' upstream region was determined by dideoxynucleotide sequencing of the mRNA. Figure 1 shows the film for one of the reactions. The film was clearly readable and, using oligonucleotides 1, 2, and 3, respectively, 111, 126, and 28 overlapping complementary nucleotide sequences were determined (Fig. 2). This information provided the sequence of 205 base-pairs upstream from the *Eco*RI site; the nucleotide sequence downstream from the *Eco*RI site (boxed) had

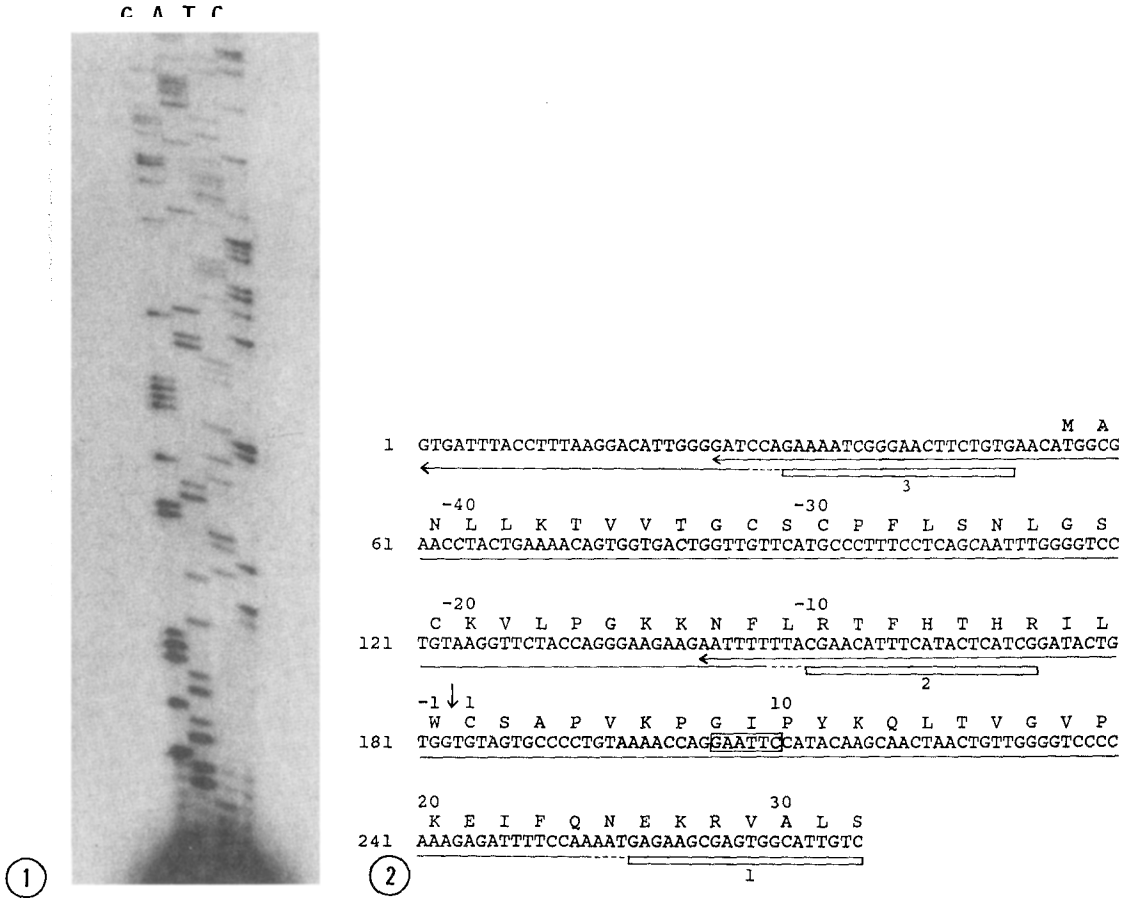


Figure 1. Autoradiogram showing the nucleotide sequence of a segment of TH mRNA. Conditions for RNA sequencing were described in "Materials and Methods". In this experiment, Oligonucleotide 2 (see Fig. 2) was used. In the autoradiogram shown, polyacrylamide gel electrophoresis was carried out for 1 h, and the nucleotide sequence of 90 bases was determined from the film. The nucleotide sequence of the additional 36 bases was determined after 2.5 h of electrophoresis.

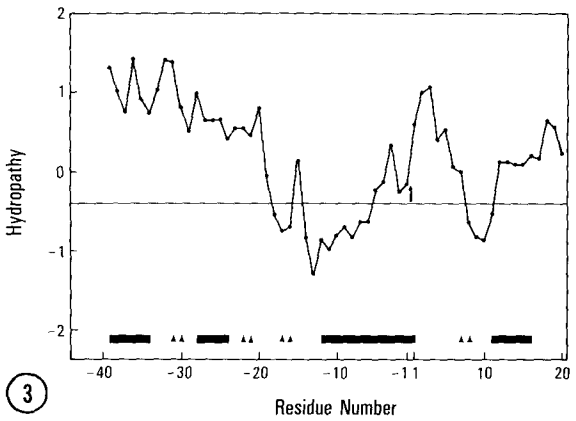
Figure 2. Nucleotide sequence encoding the N-terminal region and the signal peptide of TH. The nucleotide sequence was determined by the RNA sequencing method as described in (10). Open bars show the oligonucleotide primers (20 mers) used. The sequence of 3 base pairs (dotted lines) next to each primer was not readable on the films of experiments with that particular primer. These sequences were read with the use of other primers. The *Eco*RI site is boxed. The arrow shows where the message for the mature TH begins.

been determined previously (4). The mRNA sequence so determined encodes 24 N-terminal amino acid residues of the mature protein and 43 amino acid residues of the signal peptide. The context in which the initiation codon occurs, AACATGG, approximates the optimum (ACCATGG) (15) with the difference that C at position -2 is replaced by A, a change that does not seriously affect initiation of translation.

Further upstream, beyond the determined sequence, bands appeared on the films in two or more lanes at the same time (data not shown), suggesting premature termination of the reaction. This may suggest that the TH mRNA has some structure in the 5' upstream region or that the mRNA is not homogeneous, even though only a single band of 4.6 kb was detected in Northern blot analysis (16) using cDNA as a probe (data not shown).

N-terminal sequence of the mature enzyme - The results of protein sequencing suggested that purified TH preparations contained two TH polypeptides in a ratio of 3:2, and that the former had one extra N-terminal amino acid (4). In these preparations, the only N-terminal residue found was Ser, and the N-terminal sequence as suggested by protein sequenicng was Ser-Ser-Ala-Pro-Val-Lys-Pro-Gly ... (4). In the present study, the N-terminal sequence deduced from the mRNA sequence was Cys-Ser-Ala-Pro-Val-Lys-Pro-Gly ... (Fig. 2). Therefore, the identity of the N-terminal residue of TH was reinvestigated by manual Edman degradation (14) after carboxymethylation of the enzyme, and carboxymethylcysteine and serine were identified. Together with the mRNA data, these results indicated that the N-terminal residue of the longer polypeptide is Cys, while that of the shorter polypeptide is Ser. It is possible that the shorter chain is derived from the longer one by partial proteolytic degradation of the latter during preparation of the enzyme.

Characterization of signal peptide - From the difference between the molecular weight of the precursor protein and that of the mature enzyme, the molecular weight of the signal peptide was estimated to be 2,000 by Wu *et al.* (8) and 5,000 by Carlenor *et al.* (9). Our results agree with the latter estimate, since in the present study it was found that the molecular weight of the signal peptide as determined from the deduced amino acid sequence is 4,816. The TH signal peptide shares the following common features with other mitochondria-targeting signal peptides (17). It is rich in basic (4 Lys and 2 Arg) and hydroxylated (4 Thr and 3 Ser) amino acids, and is devoid of acidic residues. One specific feature of the TH signal peptide is that it contains 3 cysteines, which recalls the presence of disulfide-forming cysteines in membrane-permeable peptides known as defensins (18). As seen in the hydropathy plot of Fig. 3, the N-terminal half (residues -43 to -20) of the signal peptide is hydrophobic, while the C-terminal half (residues -19 to -1) is hydrophilic.



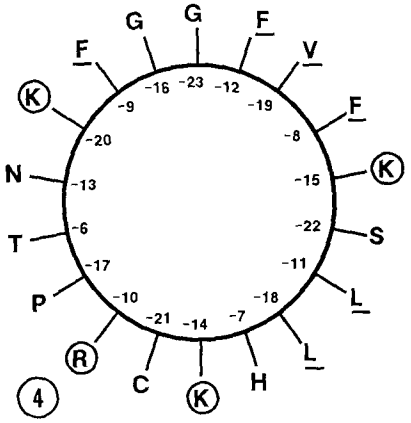


Figure 3. Hydropathy plot of the signal peptide and the N-terminal region of TH, using a setting of 9 residues. Hydropathy scores were calculated according to Kyte and Doolittle (19). The horizontal line at -0.4 on the ordinate denotes the average hydropathy of 84 fully sequenced soluble proteins. Areas above and below this line indicate relative hydrophobic and hydrophilic regions, respectively. The arrow shows where the sequence of the mature TH begins. The secondary structure was predicted by the method of Chou and Fasman (23). Bars, β-sheets; arrowheads, turns.

Figure 4. Helical wheel projection of residues -23 to -6 of the TH signal peptide. Basic residues are circled and hydrophobic residues are underlined.

According to Roise and Schatz (20), positively charged amphiphilic structure is an essential feature of mitochondrial signal peptides. When hydrophobic moments (a measure of amphiphilicity) (21) were calculated at a window setting of 18 residues (22), the segment encompassing the 18 N-terminal residue of the TH signal peptide showed a large hydrophobic moment (μ_H = 5.1), and the segment encompassing residues -23 to -6 showed an even larger hydrophobic moment (μ_H = 7.4). According to von Heijne (22), surface seeking peptides have a $\mu_H \geq 7.3$ and an $H_{max} \geq 4.5$ (H_{max} being the total hydrophobicity of 7 consecutive residues in a helical wheel projection). The H_{max} for the segment of the TH signal peptide -23 to -6 (see Fig. 4 for helical wheel projection) is 4.1. While the above calculations assume an α-helical structure, secondary structure prediction suggests that the TH signal peptide may consist mainly of β-sheets (Fig. 3). However, peptides exhibiting little secondary structure in aqueous media are known to form α-helices in membrane-like environments, and interact with the mitochondrial membranes (20,24). Furthermore, α-helicity does not appear to be a required feature of mitochondrial signal pepties, and peptides composed almost exclusively of β-sheets have been shown to penetrate phospholipid monolayers and disrupt phospholipid vesicles with great efficiency (20).

REFERENCES

1. Rydström, J. (1977) *Biochim. Biophys. Acta 463*, 155-184.
2. Fisher, R. R., and Earle, S. R. (1982) in *The Pyridine Nucleotide Coenzymes* (Everse, J., Anderson, B., and You, K.-S., eds.) pp. 279-324, Academic Press, New York.

3. Anderson, W. M., and Fisher, R. R. (1981) *Biochim. Biophys. Acta 635*, 194-199.
4. Yamaguchi, M., Hatefi, Y., Trach, K., and Hoch, J. A. (1988) *J. Biol. Chem. 263*, 2761-2767.
5. Earle, S. R., and Fisher, R. R. (1980) *J. Biol. Chem. 255*, 827-830.
6. Eytan, G. D., Persson, B., Ekebacke, A., and Rydström, J. (1987) *J. Biol. Chem. 262*, 5008-5014.
7. Wakabayashi, S., and Hatefi, Y. (1987) *Biochem. Int. 15*, 915-924.
8. Wu, L. N. Y., Lubin, I. M., and Fisher, R. R. (1985) *J. Biol. Chem. 260*, 6361-6366.
9. Carlenor, E., Joste, V., Nelson, B. D., and Rydström, J. (1985) *Bioscience Report 5*, 483-490.
10. Geliebter, J., Zeff, R. A., Welvold, R. W., and Nathenson, S. G. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3371-3375.
11. Chirgwin, J. M., Przybyla, A. E., McDonald, R. J., and Rutter, W. J. (1979) *Biochemistry 18*, 5294-5299.
12. Aviv, H., and Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1408-1412.
13. Suggs, S. V., Hirose, T., Miyake, T., Kawashima, E. H., Johnson, M. J., Itakula, K., and Wallace, R. B. (1981) *ICN - UCLA Symp. Mol. Cell. Biol. 23*, 683.
14. Chang, J.-Y. (1983) *Methods Enzymol. 91*, 455-466.
15. Kozak, M. (1984) *Nucleic Acid Res. 12*, 857-872.
16. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
17. Douglas, M. G., McCammon, M. T., and Vassarotti, A. (1986) *Microbiol. Rev. 50*, 166-178.
18. Selsted, M. E., Brown, D. M., DeLange, R. J., and Lehrer, R. I. (1983) *J. Biol. Chem. 258*, 14485-14489.
19. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105-132.
20. Roise, D., and Schatz, G. (1988) *J. Biol. Chem. 263*, 4509-4511.
21. Eisenberg, D., Schwartz, E., Komaromy, M., and Wall, R. (1984) *J. Mol. Biol. 179*, 125-142.
22. von Heijne, G. (1986) *EMBO J. 5*, 1335-1342.
23. Chou, P. Y., and Fasman, G. D. (1978) *Adv. Enzymol. 47*, 45-148.
24. Epand, R. M., Hui, S.-W., Argan, C., Gillespie, L. L., and Shore, G. C. (1986) *J. Biol. Chem. 261*,10017-10020.