Vol. 103, No. 2, 1981 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
November 30, 1981 Pages 536-542

AMINO ACID COMPOSITION AND TERMINAL SEQUENCE ANALYSIS OF THE RABIES VIRUS GLYCOPROTEIN: IDENTIFICATION OF THE READING FRAME ON THE cDNA SEQUENCE

C. Y. Lai[1] and B. Dietzschold[2]

[1]Roche Institute of Molecular Biology, Nutley, New Jersey 07110

[2]The Wistar Institute of Anatomy and Biology, Philadelphia, Pennsylvania 19104

Received October 6,1981

SUMMARY

The amino acid composition and partial sequences of amino acids at the NH_2- and COOH-termini of the rabies virus (strain ERA) glycoprotein were determined using less than 0.5 mg of the protein. The results provide identification of the reading frame on the cDNA sequence for this glycoprotein.

INTRODUCTION

The glycoprotein of rabies virus is located on the surface of the virus membrane (1), and is the only antigen of the rabies virus which reacts with and induces production of anti-rabies virus antibodies. On vaccination of mice with the antigen, the animals become protected from a challenge with rabies virus. Glycoproteins of different viral strains exhibit considerable differences in the pattern of their tryptic peptides (2). The diversity of the rabies virus glycoprotein was also demonstrated in the reactivity of different virus strains toward a series of monoclonal antibodies directed against the glycoproteins (3). In order to understand the antigenic structure and antigenic variability of the rabies virus glycoprotein, a precise chemical analysis of this protein is necessary. Recently, the nucleotide sequence of the cloned cDNA from the rabies virus glycoprotein mRNA has been determined (4). Knowledge of the amino acid composition as well as of the NH_2- and COOH-terminal sequences of the rabies virus glycoprotein has become imperative for the identification of the reading frame of the cDNA sequence.

0006-291X/81/220536-07$01.00/0

In this paper we describe for the first time the amino acid composition and partial C-terminal and N-terminal amino acid sequence of the glycoprotein of rabies virus strain ERA. Less than 0.5 mg of the purified protein was used to obtain these results.

METHODS

Preparation of Virus - Culture conditions and purification techniques have been described (5).

Purification of Rabies Virus (ERA) Glycoprotein - Purification of glycoprotein was done principally as recently described (1). Ten mg purified rabies virus was suspended in 5 ml 1% (v/v) ampholine (pH 7-9). Triton X100 was added to the suspension to a final concentration of 2%. The mixture was kept at room temperature for 20 min and centrifuged in a Beckmann SW50.1 rotor at 45,000 rpm/min for 60 min at 4°C. The supernatant was layered on a 5-40% (w/v) sucrose gradient containing 1% ampholine (pH 3.5-10) and 1% Triton X100 and electrofocused for 72 hr in an LKB 110 ml column. Fractions containing the glycoprotein were pooled. After addition of 1 volume of saturated $(NH_4)_2SO_4$ the mixture was centrifuged for 20 min at 20,000 rpm. Under these conditions the glycoprotein floats to the top of the centrifuge tube forming a pellicle. The pellicle was removed and dissolved in a small volume of distilled water. After dialysis against 0.05 M NH_4HCO_3 the glycoprotein was lyophilized and extracted twice with n-butanol and twice with ethanol to remove Triton X100. The dried material was dissolved in 1 ml 1% SDS solution and stored frozen or lyophilized. Approximately 2 mg of the pure glycoprotein, as determined by the method of S. Bramhall et al.(6), was obtained from 10 mg of the virus.

S-carboxymethylation (7) - The lyophilized sample of the glycoprotein (ERA-G) containing 3.2 mg protein and approximately 20 mg of SDS was dissolved in 1 ml of 0.3M Tris-HCl buffer, pH 8.5 containing 6M guanidine HCl, 15 mM EDTA and 10 mM DTT and incubated at room temperature for 4 hr. Twenty µl of 1 M iodoacetic acid was added and the incubation continued for 30 min. After addition of 1 drop of mercaptoethanol, the reaction mixture was dialyzed against 500 ml of 10 mM NH_4HCO_3, 3mM NaN_3 in the cold room, and stored frozen.

Amino acid analysis - This was performed either a with JEOL-6AH automatic amino acid analyzer (JEOL Inc., Cranford, N.J.) or a single-column analyzer equipped with a fluorescamine detection system (8). The levels of sensitivity are 1 nmol with the former and 50 pmol with the latter instrument.

Sequence analyses - The NH_2-terminal sequence was determined by Edman degradation of S-carboxymethyl ERA-G essentially as described by Tarr (9), with the following modifications: The dialyzed sample (0.1 ml containing 120 µg S-carboxymethyl ERA-G) was mixed with 15 µl of 10 mg/ml polybrene and about 20 mg prewashed glass beads in a 1 ml Reacti-Vial (Pierce Chemical, Rockford, Ill.) and lyophilized. The dried material was suspended in 30 µl of the coupling buffer (1:1 mixture of 25% trimethylamine-pyridine), lyophilized and then subjected to Edman degradation. The coupling and cleavage reactions were carried out at 50° for 20 min and 5 min twice, respectively. The anilinothiazolinone-amino acid extracted at each step was hydrolyzed in 0.1 ml of 0.1% $SnCl_2$ in 5.7 N HCl at 150° for 4 h to regenerate the amino acid for identification with the amino acid analyzer (10).

The COOH-terminal sequence was analyzed by following the time-course release of amino acid from ERA-G after digestion with carboxypeptidases A and B (DFP-treated, Worthington Biochemical, Freehold, N.J.).

The COOH-terminal residue was also determined by hydrazinolysis of S-carboxymethyl ERA-G (11). The sample (60 µg) was heated with 0.1 ml anhydrous hydrazine (Eastman Organic Chemicals) in a evacuated tube for 5 or 15 h at 105°. After drying, it was dissolved in 0.1 ml H_2O, treated with 10 µl isovarylaldehyde for 1 h at room temperature, and the mixture extracted with 0.2 ml ether 3 times. The aqueous layer containing the freed COOH-terminal residue was then analyzed.

RESULTS

The purified ERA-G is soluble only in the presence of 1% SDS at an elevated temperature after boiling. Aliquots were withdrawn for analyses from a hot solution. On reduction and S-carboxymethylation, the material remained in solution after dialysis (See "Methods") but formed a precipitate during repeated freezing and thawing.

Amino acid composition of ERA-G - The dry weight of the sample in the hydrolyzate was calculated from the weights of all amino acids except tryptophan as determined by the amino acid analysis. From this value, the number of moles of ERA-G was calculated based on M_r = 60,000 (1). The amino acid composition of ERA-G (Table 1) is in excellent agreement with that deduced from the cDNA sequence after identification of the reading frame (4).

The NH_2-terminal sequence of ERA-G - the results of Edman degradation indicate that the sequence of 6 amino acid residues at the NH_2-terminus of ERA-G is:

(NH_2) Lys-Phe-Pro-Ile-Tyr-Thr- (Table 2)

The COOH-terminal sequence - The COOH-terminal residue was identified as leucine by the hydrazinolysis procedure: Leucine was recovered in 42.7% yield from 90 µg (1.5 nmol) of the protein after 5 h of hydrazinolysis at 105°. Virtually no other amino acids were found.

The result of carboxypeptidase digestion unequivocally established the penultimate amino acid as arginine (Table 3).

Table I Amino acid composition of ERA-G

	Residues per mole	
	Amino acid[a] analyses	Predicted from cDNA sequence (4)
Lysine	28	28
Histidine	18	18
Arginine	28	28
Aspartic Acid	51	46
Threonine	31	34
Serine	39	42
Glutamic Acid	43	39
Proline	30	28
Glycine	39	38
Alanine	21	20
Half Cystine	13	16
Valine	38	39
Methionine	13	13
Isoleucine	23	22
Leucine	50	51
Tyrosine	16	16
Phenylalanine	19	18
Tryptophan	-	9

a) S-carboxymethyl ERA-G (40 µg) was hydrolyzed with 5.7 N HCl at 110° for 22, 48 and 72 hr and the hydrolyzate analyzed with a Jeol 6AH amino acid analyzer. The values shown are the average of the 3 analyes, except for Ser and Cys where the values from the 22 hr hydrolyzate are shown. Cysteine was determined as S-carboxymethyl cysteine.

DISCUSSION

With availability of a high-sensitivity amino acid analyzer and the method for quantitative regeneration of amino acids from their AT-derivatives (10), it has been possible to perform Edman degradation with 1-2 nmol (50-100 µg of the ERA strain glycoprotein) of proteins in this laboratory[1]. The rabies virus glycoprotein is practically insoluble in water in the absence of detergents. After S-carboxymethylation and extensive dialysis, the protein remained in solution but formed a membranous precipitate on storage at 3° C. Addition of glass beads and polybrene to the sample during Edman degradation was to insure a good

Table II Edman degradation of ERA-G

Steps	Amino acid found (%)[a]	Residue identified
1	Lys (17), Gly (10)	Lys[b]
2	Phe (50), Pro (17), Lys (8)	Phe
3	Pro (92) αAB (4)	Pro
4	Ile (60)[c], Tyr (7), αAB (4)	Ile
5	Tyr (19), αAB (11)[c]	Tyr
6	αAB (22), Leu (7)	Thr

NH_2-terminal sequence: Lys-Phe-Pro-Ile-Tyr-Thr-

a) Edman degradation was performed on 120 µg (2 nmole) of ERA-G. Amino acids found in each step are shown with % yield in parenthesis (10). Similar results were obtained in another experiment with lower yields.
b) Also identified as its PTH-derivative by reverse phase chromatography.
c) The sum of alloisoleucine and isoleucine (10).
αAB=α-amino-butyric acid

dispersion of the glycoprotein particles at each step of the reaction. Despite the insolubility of the material, over 70% repetitive yield was obtained in the degradation steps indicating the high efficiency of the procedure.

The complete nucleotide sequence of the cDNA complementary to mRNA of the ERA strain glycoprotein has recently been determined. One reading

Table III Carboxypeptidase digestion of ERA-G[a]

	Treatment at 23°	Amino acid found ($\frac{\text{residue}}{\text{mol ERA-G}}$)
Exp. 1	CpA added at 0 min.	At 15 min; Leu (0.46) 2 hr; Leu (0.5)
Exp. 2[b]	CpA added, boiled after 15 min.	Leu (0.91)
	CpB added to above, boiled after 15 min.	Leu (1.0), Arg (0 79)

a) A suspension of ERA-G (50 µg 0.8 nmol) in 0.05 M NH_4HCO_3 containing 1.2% SDS (3µg/µl) was diluted with 2 vol of H_2O and treated with carboxypeptidase A (CpA)(1.6 µg) (CpA) or carboxypeptidase B (CpB) (4 µg) as indicated. At indicated time an aliquot was removed, and analyzed for amino acid released.
b) Same as Exp. 1 except ERA-G suspension was boiled before addition of CpA. Further addition of CpA resulted in the release of many amino acids in 3 min, making the time course analysis impossible.

frame starting with the initiation codon ATG, located at nucleotides 8 - 10 from the 5' end of cDNA, was found to code for a polypeptide of 524 residues without encountering a stop codon TGA (4). The search was also facilitated by the knowledge that the protein had Leu at the COOH-terminus. The NH_2-terminal sequence of the ERA strain glycoprotein determined here, (NH_2) Lys-Pre-Pro-Ile-Tyr-Thr (Table 2), was found to match residues 20 to 25 of the polypeptide predicted from the above cDNA sequence, and thus established it as the correct reading frame. The COOH-terminal sequence, -Arg-Leu (OH) (Table 3), was found to occur only at residues 523 - 524 of the deduced polypeptide, and the nucleotide sequence was followed by the stop codon TGA. Since the apparent molecular weight of ERA-G has been estimated experimentally as 60,000, this was in excellent agreement with the length of the polypeptide deduced from the cDNA sequence. Moreover, the amino acid composition of the predicted polypeptide from this reading frame agreed well with that determined in this report (Table 1). The discrepancies are within 10% which is the range of error expected from the amino acid analysis.

The amino acid and terminal analyses of ERA-G in conjunction with determination of its cDNA sequence provides an example for a new and powerful approach to elucidation of the primary structure of a protein. Structural variations in the glycoproteins from different rabies virus strains may eventually be elucidated using a similar approach.

REFERENCES

1. Dietzshold, B., Cox, J.H., Schneider, L.G., Wiktor, T.J. and Koprowski, H., J. Gen. Virol. 40 131 139 (1978).
2. Dietzshold, B., Cox, J.H. and Schneider, L.G., Virology 98 63-75 (1979).
3. Wiktor, T.J., Flamand, A. and Koprowski, H., J. Virol. Meth., 1 33-36 (1980).
4. Anilionis, A., Wunner, W.H., and Curtis, P.J., Nature, submitted.
5. Wiktor, T.J.,Dietzshold, B., Leamnson, R.N. and Koprowski, H., J. Virol. 21 626-635 (1977).
6. Bramhall S., Noack, N., Wu, M., and Loewenberg, J.R.,Anal. Biochem. 31 146-148 (1969).

7. Crestfield, A.M., Moore, S. and Stein, W.H., J. Biol. Chem. 238 622-627 (1963).
8. Stein,S., Bohlen, P., Stone,J., Dairman, W. and Udenfriend, S., Arch. Biochem. Biophys. 155 202-210 (1973).
9. Tarr, G.E., in Methods in Enzymology 47 (Hirs and Timasheff eds.) 335-357 (1977).
10. Lai, C.Y., in Methods in Enzymology 47 (Hirs and Timasheff eds.) 369-373 (1977).
11. Niu, C.I. and Fraenkel-Conrat, J. Am. Chem. Soc. 77 5882-5884 (1955).