Localization of Translation Initiation in the Message for E. coli UDP-Galactose 4-Epimerase: The Amino Terminal Sequence

D.H. Schlesinger, H.D. Niall and D. Wilson*

Endocrine Unit, Massachusetts General Hospital, Boston, Massachusetts, 02114 and Department of Biochemistry*, Cornell University, Ithaca, New York, 14850.

Received September 18,1974

SUMMARY

The amino acid sequence of E. coli UDP-galactose 4-epimerase has been determined through the amino-terminal 28-amino acid residues using an automated protein sequenator. Alignment of UDP-galactose operon messenger RNA and the amino acid sequence of epimerase demonstrates that the first 26 bases in the mRNA are transcribed but do not take part in translation of epimerase.

INTRODUCTION

UDP-galactose 4-epimerase has been the subject of many studies aimed at determining the mechanism by which it catalyzes the interconversion of UDP-glucose and UDP-galactose (1,2). In recent years these studies have mainly utilized the enzymes from E. coli and yeast (3,4).

UDP-galactose 4-epimerase is an exception to the general rule that enzymes involved in catabolic pathways do not catalyze biosynthetic reactions. The reaction catalyzed by this enzyme is the sole pathway for the biosynthesis of UDP-galactose in cells growing in the absence of galactose

and is also an essential step in the conversion of galactose to glucose 1-phosphate, the catabolic pathway that allows E. coli to utilize galactose as a carbon source (5). The structural gene for this enzyme maps in the galactose operon and is adjacent to the operator gene (6). The regulation of this system has been the subject of may studies (7--15), and currently de Crombrugghe is determining the sequence of E. coli galactose operon messenger RNA. Because the structural gene for this enzyme is at the operator end of the galactose operon and because transcription and translation both start at this end of the operon, a knowledge of the NH_2-terminal sequence of E. coli UDP-galactose 4-epimerase allows the determination of the first translation-initiation signal in the galactose mRNA, and the sequence between this site and the 5' end should include the galactose operator sequence. In this study we have used a sequenator to determine the amino acid sequence of the first 28 residues starting with the amino terminal residue of pure E. coli UDP-galactose 4-epimerase.

EXPERIMENT

The strain of bacteria used was E. coli K12 having the genotype c^-, ade^-, thi^-, gal T^-, gal R^c lysogenic for λQSdg from which the gal operon or the phage was derived from p gal_8. This strain was grown in a 12-liter fermenter (New Brunswick Scientific) in 10 liters of media containing 16 grams Bacto-tryptone, 10 grams yeast extract, and 5 grams sodium chloride per liter. The cells were grown at 30°C to a density of 7 x 10^8/ml, and then the temperature was raised to 45°C, kept there for three minutes, and then dropped to 37°C. Growth was continued for an additional two hours, and then the

temperature was reduced and the cells harvested as quickly as possible using a refrigerated Sharples continuous-flow centrifuge. The cell pellet was stored frozen at -20°c.

The UDP-glucose dehydrogenase used in the epimerase assay (3) was fraction V from the purification of Wilson (16). Protein was determined by the Lowry procedure (17).

UDP-galactose 4-epimerase was purifed as described previously (3) except that the frozen cells (16 grams) were thawed, homogenized in 300 ml of 0.01 M KPO_4 (pH 7.0) containing 10^{-3} M EDTA. This treatment was sufficient to lyse the cells and prepare the extract. The only other modification was that the hydroxylapatite (Hypatite C, Clarkson Chem. Co.) column was run at pH 7 instead of 6.5, after the DEAE sephadex A50 column instead of before it. The purity of epimerase was evaluated by electrophoresis on a 7 1/2 % polyacrylamide gel using the standard Tris--glycine discontinuous buffer system (18), as well as NH_2 -terminal analysis (20).

Performic acid oxidation of UDP-galactose 4-epimerase was performed in the following manner. The enzyme (4 mg) was dissolved in 300 µl of performic acid prepared by mixing 9.5 ml formic acid (88%) and 1.0 ml hydrogen peroxide (30%) at 0°C. Oxidation proceeded for 2 hr at 0°C, after which the protein was diluted with distilled water and lyophylized. After drying, dilution and lyophilization were repeated to remove traces of acid.

Native and performic-acid-oxidized UDP-galactose 4-epimerase was subjected to automated sequential Edman degradations on a sequenator (Beckman model 890) using 1 M Quadrol buffer according to a modified procedure of

Edman and Begg (19). (All reagents used in sequential analysis were purchased from Beckman Instruments) The thiazolinone resulting at each step of the degradation was converted to its isomeric phenylthiohydantoin amino acid by heating in 0.2 ml of 1.0 N HCl at 80° for 10 min. (20).

PTH amino acids residing in the ethylacetate extract after conversion were identified by thin-layer chromatography using system H_1 of Edman (21). Confirmation of these identifications was made by gas chromatography (Beckman GC 45 gas chromatograph) under isothermal conditions with Beckman SP400 as the solid-support resin. The polar PTH amino acids, arginine, histidine, and cysteic acid, remaining in the aqueous phase after conversion were identified by thin-layer chromatography using a solvent composed of chloroform: methanol: heptafluorobutyric acid, 70:30:0.5. Further identification and confirmation of these polar PTH amino acids were accomplished by gradient liquid chromatography (Waters liquid chromatograph ALC 202) using a C_{18} microporasil column and a modified procedure of Zimmerman et al.(22).

RESULTS

The purified UDP-galactose 4-epimerase had a specific activity of 15,000 units/mg, showed a single band when subjected to "disc" electrophoresis, and released only one PTH amino acid (PTH methionine) on end-group analysis. This native enzyme was subjected to automated Edman degradation. The sequence elucidated by this analysis is presented in Fig. 1 which depicts the yield of the PTH amino acid at selected steps of the degradation. Positive identifications were made at each step of the degradation through residue

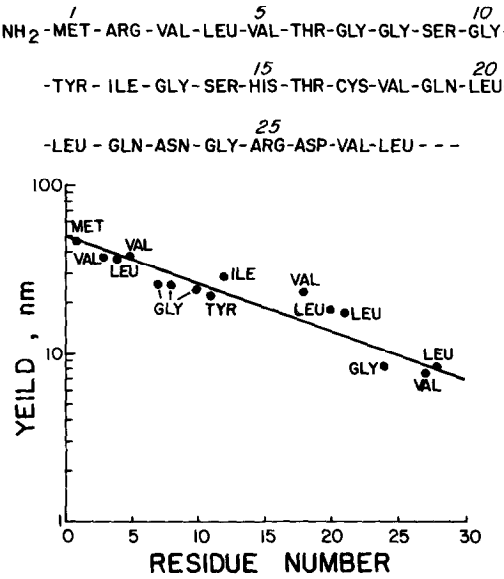

Fig. 1. Repetitive-yield curve of PTH amino acids obtained at selected steps of the automated degradation of native UDP-galactose 4-epimerase. Background PTH amino acid was subtracted to give actual PTH yield. The complete amino acid sequence of the first 28 residues of UDP-galactose 4-epimerase is presented at the top of the figure.

---

28 with the exception of Step 2 (which was tentatively identified as arginine) and Step 17, which remained unidentified in this degradation. The calculated average yield per cycle from Step 3 (valine) to Step 27 (valine) or Step 4 (leucine) to Step 28 (leucine) was found to be 94.2% and 94.6% respectively.

In order to identify the residue released at Step 17 and to reconfirm the residue at Step 2, as well as the other steps in this initial degradation, E. coli epimerase was performic-acid oxidized and then degraded by the automated Edman procedure. Results of this second degradation confirmed the assignment of arginine at Step 2, as well as the other assignments, and identified position 17, which was found to be PTH-cysteic acid. The results of

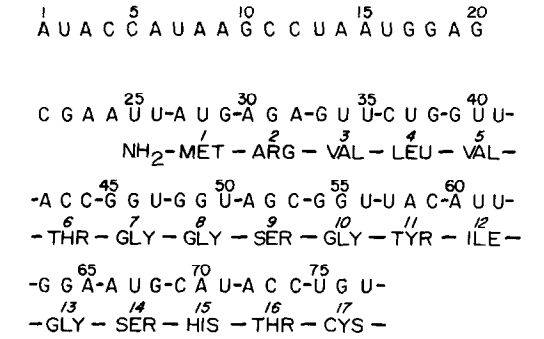

Fig. 2. Correlation of Gal mRNA and UDP-galactose 4-epimerase sequences.

these two degradations completes the NH_2-terminal amino acid sequence of the first 28 residues of UDP-galactose 4-epimerase.

DISCUSSION

This study confirms an earlier investigation (23) reporting that the two subunits of epimerase have an identical NH_2-terminal amino acid sequence. The NH_2-terminal amino acid is methionine, however, and not aspartic acid as was reported (23) in the earlier study.

The structural gene for UDP-galactose 4-epimerase maps at the operator end of the galactose operon. Since transcription and translation begin at this end of the operon, the N-terminal sequence of epimerase allows the determination of the first translation-initiation signal in the galactose operon mRNA. Recently, the partial sequence of E. coli galactose operon messenger RNA has been elucidated (24). Alignment of this mRNA base sequence and the amino acid sequence of epimerase (Fig. 2) demonstrates that the first 26 bases in mRNA do not take part in translation of epimerase. Protein-chain initiation begins with

AUG (in positions 27--29 in the mRNA base sequence) which codes for the NH_2-terminal methionine in epimerase. Starting with residue 27, the amino acid sequence of epimerase exactly corresponds to the mRNA sequence through position 77. The base sequence beyond 77 has not yet been established, but the amino acid sequence in positions 18--28 that we report here should be helpful in predicting the mRNA nucleotide sequence within the limits imposed by the degeneracy of the code.

ACKNOWLEDGEMENT

We wish to thank Dr. Ira Pastan for arranging collaboration between laboratories of the Endocrine Unit at the Massachusetts General Hospital and the Deparmtent of Biochemistry at Cornell University and to Mr. Phillip Dee of the Endocrine Unit for the illustrations.

This work was supported by grants AM 16168, AM 11794, and 5 RO1AM11923 from the National Institutes of Health.

REFERENCES

1. Rose, I.A., Ann. Rev. Biochem. 35,23 (1966).
2. Glaser, L., In: The Enzymes. 3d edition. Edited by P.D. Boyer, Vol. VI, p. 355 (1972).
3. Wilson, D.B., and Hogness, D.S., J. Biol. Chem. 239, 2469 (1964).
4. Darrow, R.A., and Rodstrom, R. Biochemistry 7,1645 (1968).
5. Kalckar, H.M., Advances Enzymol. 20,111 (1958).
6. Adler, J., and Kaiser, A.D., Virology 19, 117 (1963).
7. Jordon, E., Yarmolinsky, M., and Kalckar, H.M., Proc. Nat. Acad. Sci. USA.48,32 (1962) .
8. Buttin, G., J. Mol. Biol. 7,164 (1963).

9. Adhya, S., and Echols, H., J. Bact. 92,601 (1966).

10. Michaelis, G., and Starlinger, P. Mol. Gen. Genet. 100, 210 (1967).

11. Wilson, D.B., and Hogness, D.S., J. Biol. Chem. 244,2143 (1969).

12. Mackie, G., and Wilson, D.B., J. Biol. Chem. 247,2973 (1972).

13. Hua, and Markovitz, A., J. Bact. 110,1089 (1972).

14. Nakanishi, S., Adhya, S., Gottesman, M.E. and Pastan, I., Proc. Nat. Acad. Sci. USA 70,334 (1973).

15. Schumacher, G., and Ehring, R., Mol. Gen. Genet. 124, 329 (1973).

16. Wilson, D.B., Anal. Biochem. 10,472-487 (1965).

17. Lowry, O.H., Rosebrough, N.J., Farr, A.L., and Randall, R.S., J. Biol. Chem. 193,265 (1951).

18. Davis, B.S., Ann. N.Y. Acad. Sci. 121,404 (1964).

19. Edman, P. and Begg, G., Europ. J. Biochem 1,80 (1967).

20. Niall, H.D., Methods Endocrinol. 27, 942 (1973).

21. Edman, P., In: Protein Sequence Determination. Edited by S. Needleman, p. 211, Springer, Verlag (1970).

22. Zimmerman, C., Pizano, J., and Appella, E., Biochem. Biophys. Res. Commun. 5,1220(19 73).

23. Wilson, D.B., and Hogness, D.S., J. Biol. Chem., 244,2132 (1969).

24. Musso, R.E., de Crombrugghe, B., Pastan, I., Skylar, J., Yot, P., and Weissman, S., Proc. Nat. Acad. Sci. USA (in press).