# Isolation of a fourth cysteinyl-containing peptide of the α-subunit of the $F_1$ ATPase from *Escherichia coli* necessitates revision of the DNA sequence

Helga Stan-Lotter, David M. Clarke and Philip D. Bragg

*Department of Biochemistry, University of British Columbia, Vancouver, BC, Canada*

Received 20 December 1985

The rapid determination of cysteinyl residues by Creighton's method [(1980) Nature 284, 487–489] led to the discovery of a discrepancy between protein and DNA sequence data in the α-subunit of the $F_1$ ATPase from *Escherichia coli* [(1984) Arch. Biochem. Biophys. 229, 320–328]. We have isolated a cysteinyl-containing decapeptide from the α-subunit with a protein sequence (AGCAMGEYFR) which is only partially recognizable from DNA data. Re-sequencing of DNA in the region coding for the peptide has resulted in two corrections: insertion of a cytosine before position 715 and deletion of a thymine at position 731 of the *uncA* gene.

*Amino acid sequence* *DNA sequence* *Cysteinyl residue* *$F_1$-ATPase* (E. coli)

## 1. INTRODUCTION

Earlier we reported the determination of the number of cysteinyl residues of the α-subunits of the $F_1$ ATPases from *Escherichia coli* and *Salmonella typhimurium* by various protein chemical methods [1,2]. In both organisms we found 4 cysteinyl residues in the α-subunit. The most unambiguous data were obtained by using Creighton's method [3]. This result was at variance with the data from DNA sequencing of the $F_1$ ATPase genes of *E. coli* [4,5], which predicted 3 cysteinyl residues for the α-subunit. Initially we supposed that an unusual post-translational modification, to add a cysteinyl residue to the α-subunit, might have occurred. This assumption could not be verified, as is shown here. We then proceeded to isolate cysteinyl-containing peptides from the α-subunit. One decapeptide partially differed in its amino acid sequence from that predicted by the DNA sequence. We have re-sequenced the DNA of the α-subunit which codes for this region. Two corrections of the published DNA sequences [4,5] have to be made, one insertion and one deletion. The resulting amino acid sequence, which includes the decapeptide and two residues preceding it, is in good homology to corresponding portions of the α-subunit of 5 other $F_1$ ATPases from different organisms [6].

*Abbreviations:* IEF, isoelectric focusing; HPLC, high-pressure liquid chromatography; DPCC trypsin, diphenylcarbamyl chloride-treated trypsin

## 2. MATERIALS AND METHODS

### 2.1. *Preparation of ATPase and α-subunit*

The ATPase of *E. coli* strain KY7485 and its α-subunit were prepared as described [1,7,8].

### 2.2. *Determination of thiol groups of in vitro synthesized α-subunit*

DNA from the plasmid pRPG54, which contains the entire *unc* operon [9], was prepared according to [10]. The components for the coupled transcription-translation system were prepared as described [11]. The S30 extract was prepared from *E. coli* strain A19 or AB1157 (wild type). In vitro

*Published by Elsevier Science Publishers B.V. (Biomedical Division)*
00145793/86/$3.50 

synthesis was carried out for 45 min according to [11]. The final DNA concentration was 0.047 mg/ml. Proteins were labeled by using [$^{35}$S]methionine (1000 Ci/mmol, Amersham). The synthesis mixture was alkylated with iodoacetamide or iodoacetate and the number of thiol groups determined according to Creighton [3] with the modification described in [12].

### 2.3. *Isolation and sequencing of thiol-containing tryptic peptides of the α-subunit*

14 mg purified α-subunit in 50 mM Tris-HCl, pH 8, 1 mM EDTA, 2 M urea, 0.1 mM DTT, was immobilized in a slurry of 2 g thiopropyl-Sepharose 6B (Pharmacia) and digested for 22 h with DPCC-trypsin (Sigma) following the procedure in [13] with minor modifications. Thiol-containing peptides were eluted with 50 mM 2-mercaptoethanol in 50 mM ammonium bicarbonate. The recovery was about 70%. Prior to fractionation by HPLC, the peptides were alkylated with iodoacetamide or iodoacetic acid in the presence of 8 M urea. The mixture was immediately applied to a Dupont Zorbax C8 column, 9.4 mm × 25 cm, and excess reagents removed by washing with 0.05% trifluoroacetic acid. Peptides were separated by applying a gradient of 0–80% acetonitrile in 0.05% trifluoroacetic acid (v/v). The amino acid sequences were determined by Edman degradation on a gas phase sequencer (Applied Biosystems 470A).

### 2.4. *Sequencing of DNA*

Plasmid pRPG54 was treated with the restriction endonucleases *Eco*RI and *Bgl*II (Pharmacia). The *Eco*RI/*Bgl*II fragments were cloned into the *Eco*RI and *Bam*HI sites of M13 mp18. Phage containing the desired 806 base-pair (bp) fragment from the *uncA* region was selected and single-stranded M13 templates for DNA sequencing were prepared under standard conditions. The DNA sequence was determined by the dideoxy chain termination method [14] using a 17-nucleotide long primer complementary to a region adjacent to the linker sequence of M13 mp18.

## 3. RESULTS AND DISCUSSION

### 3.1. *In vitro synthesized α-subunit contains 4 thiol groups*

The discrepancy between the number of cysteinyl residues of the α-subunit from the *E. coli* $F_1$ ATPase as determined by protein chemical methods [1] or DNA sequencing [4,5] prompted us to examine the protein after in vitro synthesis. If the DNA sequence was correct, the extra cysteinyl residue found in the protein might have been added as a post-translational modification. $F_1$ ATPase was synthesized in vitro using the DNA of plasmid pRPG54 [9] as a template (fig.1, lane 1). The thiol groups of the α-subunit were conveniently determined by the method of Creighton [3], which utilizes the incorporation of one negative charge per cysteinyl residue into the protein following carboxymethylation. To avoid overlapping with plasmid proteins, the modified α-subunits were excised from SDS gels before separation on IEF gels [12]. The α-subunit formed by in vitro synthesis appeared as two species on an IEF gel, differing in one charge (fig.1, lane 2). Both species incorporated 4 negative charges following treatment with iodoacetate, indicating the presence of 4 cysteinyl residues (fig.1, lanes 3–7). A similar result was obtained when the in vitro synthesis was carried out for only 10 min instead of 45 min. These results show that a post-translational modification by addition of a cysteinyl residue is unlikely.

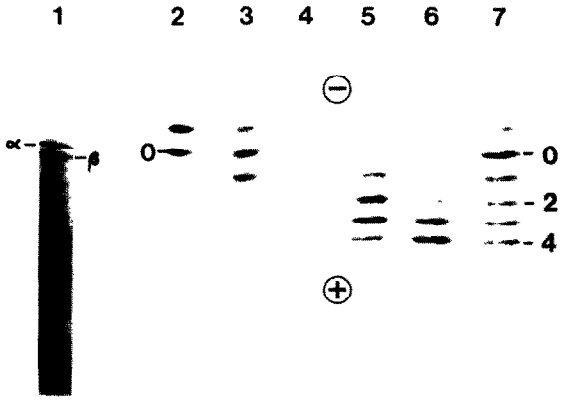


Fig.1. Cysteinyl residues of the α-subunit formed by in vitro synthesis. Lane 1: SDS gel electrophoresis of in vitro protein synthesis products of plasmid pRPG54. The major subunits of $F_1$ ATPase are indicated. Lanes 2–7: isoelectric focusing gel following SDS gel electrophoresis of α-subunit excised from gels and alkylated with iodoacetamide (iam) and/or iodoacetate (iac). The figures on the right show the number of charges introduced by reaction of thiol groups with iodoacetate. Lanes: 2, iam; 3, iam:iac = 1:1; 4, iam:iac = 1:3; 5, iam:iac = 1:9; 6, iac; 7, mixture of samples 2–6.

### 3.2. *Isolation of thiol-containing tryptic peptides from the α-subunit*

Thiol-containing peptides of the α-subunit were isolated by affinity chromatography on thiopropyl-Sepharose 6B. Subsequent separation of peptides was performed by HPLC. Based on the DNA sequence of the α-subunit [4,5], tryptic cysteinyl-containing peptides comprise amino acid residues 41–59 (named TT1), 90–93 (TT2) and 193–201 (TT3). Three peptides were eluted at 43.5, 45.5 and 69% acetonitrile, respectively. The amino acid analysis showed unambiguously that peptide TT3 eluted at 43.5% acetonitrile and peptide TT1 at 69% acetonitrile. The amino acid sequence of peptide TT3 was determined and agreed with the predictions from the DNA sequences [4,5]. The smallest thiol peptide, TT2, was not found. It is probable that it did not bind to the reverse-phase resin, as is known for some small peptides [15], and therefore was lost during the washing step. The peptide eluting at 45.5% acetonitrile was called TT4. Its amino acid composition did not agree with any of the tryptic peptides of the α-subunit as predicted from the DNA sequence [4,5].

### 3.3. *Amino acid sequence of peptide TT4 and re-sequencing of the DNA from the* uncA *gene*

The amino acid sequence of peptide TT4 as determined by Edman degradation is shown in fig.2, line 3. The last 6 residues of the peptide were identified as amino acids 245–250 of the α-subunit [4,5]. The sequence of the first 4 residues, Ala-Gly-Cys-Ala, does not agree with that predicted from the DNA sequence (fig.2, line 1).

We chose the fragment between the *Eco*RI and the *Bgl*II sites of the *uncA* gene in order to redetermine the DNA sequence around the region of discrepancy. This fragment consists of 806 bp, 127 of which were sequenced. The DNA sequence between nucleotides 715 and 750 is shown in fig.2, line 4. There is a difference at two positions when compared with the DNA sequence shown in fig.2, line 2: one cytosine is inserted before position 715 and one thymine is deleted at position 731. The remainder of the DNA sequence, extending to nucleotides 689 and 815, respectively, agrees completely with published sequence data [4,5] (not shown).

The amino acid sequence as deduced from the

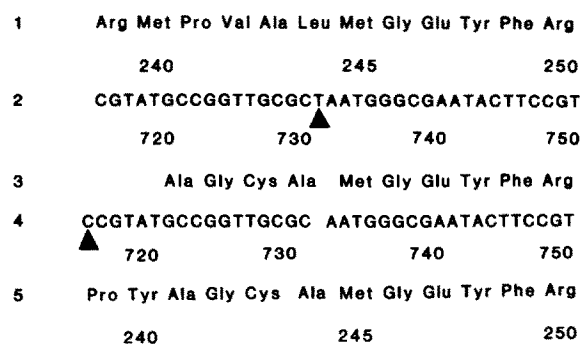


Fig.2. Correction of amino acid and DNA sequences of the α-subunit between bases 715 and 731. Lines: 1, amino acid sequence of residues 239–250 of the α-subunit as deduced from the DNA sequence below; 2, DNA sequence of nucleotides 715–750 of the α-subunit according to [4,5]; 3, amino acid sequence of peptide TT4, which was isolated as described in this paper; 4, corrected DNA sequence of nucleotides 715–750 of the α-subunit; 5, amino acid sequence of residues 239–250 of the α-subunit as deduced from the DNA sequence in line 4.

corrected DNA sequence is shown in fig.2, line 5. It contains the sequence of peptide TT4 as well as proline at position 239 and tyrosine at position 240. Pro-239 (*E. coli* numbering) of the α-subunit is conserved in all $F_1$ ATPases which have been investigated so far [6]. Other sequence homologies include Tyr-240, Gly-242, Cys-243 and Ala-244, which are conserved in several of the α-subunits from bacterial, plant and eukaryotic $F_1$ ATPases [6].

There is no trypsin cleavage site at position 240. It is likely, however, that residual chymotryptic activity in DPCC-trypsin gave rise to peptide TT4.

## 4. CONCLUSIONS

The correctness of long DNA sequences is usually established by determining short segments of the N-terminal and C-terminal sequences of the protein. However, incorrect segments in the center will escape detection unless stop codons occur. The fast and reliable determination of the number of cysteinyl residues in a protein [3,12] is a method which can lead to the identification of incorrect DNA sequences, as is demonstrated in this report.

## ACKNOWLEDGEMENTS

We thank Sandy Kielland (Protein Sequencing Facility, University of Victoria) for determining the amino acid sequences, Dr Marcia Mauk for helpful advice with the HPLC system and Dr Robert Simoni for providing the plasmid-containing strain. This work was supported by the Medical Research Council of Canada.

## REFERENCES

[1] Stan-Lotter, H. and Bragg, P.D. (1984) Arch. Biochem. Biophys. 229, 320–328.
[2] Stan-Lotter, H. and Bragg, P.D. (1985) Arch. Biochem. Biophys. 239, 280–285.
[3] Creighton, T.E. (1980) Nature 284, 487–489.
[4] Kanazawa, H., Kayano, T., Mabuchi, K. and Futai, M. (1981) Biochem. Biophys. Res. Commun. 103, 604–612.
[5] Gay, N.J. and Walker, J.E. (1981) Nucleic Acids Res. 9, 2187–2194.
[6] Walker, J.E., Fearnley, J.M., Gay, N.J., Gibson, B.W., Northrop, F.D., Powell, S.J., Runswick, M.J., Saraste, M. and Tybulewicz, V.L.J. (1985) J. Mol. Biol. 184, 677–701.
[7] Dunn, S.D. and Futai, M. (1980) J. Biol. Chem. 255, 113–118.
[8] Bragg, P.D., Stan-Lotter, H. and Hou, C. (1982) Arch. Biochem. Biophys. 213, 669–679.
[9] Gunsalus, R.P., Brusilow, W.S.A. and Simoni, R.D. (1982) Proc. Natl. Acad. Sci. USA 79, 320–324.
[10] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[11] Brusilow, W.S.A., Gunsalus, R.P. and Simoni, R.D. (1983) Methods Enzymol. 97, 188–195.
[12] Stan-Lotter, H. and Bragg, P.D. (1985) Can. J. Biochem. Cell Biol., in press.
[13] Rydén, L. and Norder, H. (1981) J. Chromatogr. 215, 341–350.
[14] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[15] Slikowski, M.X. and Levine, R.L. (1985) Anal. Biochem. 147, 369–373.