Vol. 207, No. 1, 1995 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
February 6, 1995 Pages 331-338

GLUCOSE 6-PHOSPHATE DEHYDROGENASE FROM HUMAN ERYTHROCYTES: IDENTIFICATION OF N-ACETYL-ALANINE AT THE N-TERMINUS OF THE MATURE PROTEIN

Laura Camardella[1*], Gianluca Damonte[2], Vito Carratore[1], Umberto Benatti[2], Michela Tonetti[2], and Gloriano Moneti[3]

[1]Institute of Protein Biochemistry and Enzymology, C.N.R., Napoli, Italy

[2]Institute of Biochemistry, University of Genova and Advanced Biotechnology Center, Genova, Italy

[3]Department of Pharmacology, University of Firenze, Firenze, Italy

Received December 20, 1994

Glucose 6-phosphate dehydrogenase from human erythrocytes has a blocked amino-terminus and no information could be obtained by direct sequencing of the intact protein. The peptide corresponding to the amino-terminal region was isolated from a tryptic digest of the whole protein and identified on the basis of its amino acid composition and of the failure to obtain Edman degradation. Determination of peptide mass by fast atom bombardment mass spectrometry allowed identification of the blocked amino-terminal residue as N-acetyl-alanine.

Glucose 6-phosphate dehydrogenase (EC 1.1.1.49, G6PD), the enzyme which catalyzes the first step of the pentose phosphate pathway, has attracted the interest of many researchers because of its importance as a target of metabolic regulation (1) and also from the standpoint of human biochemical genetics (2). It is a ubiquitous enzyme, and is present in human red blood cells, where it is mainly involved in the protection of the cell integrity against oxidation by maintaining the required level of reduced NADP.

Since the discovery that a genetically inherited enzyme deficiency is associated with a maternal transmission, a large number of X-linked G6PD variants has been studied (3).

The enzyme is in a tetramer-dimer equilibrium affected by environmental factors (1). Its identical subunits have an approximate molecular mass of 58 kDa (4). The primary structure has been determined both by direct amino acid sequencing (5) and by extensive sequencing of human cDNA clones (6). The DNA-deduced amino acid sequence is composed of 515 residues and contains the correct C-terminal sequence (7). A major discrepancy in the N- terminal region between the sequence deduced from cDNA and

*Corresponding author.Fax: +39.81.5936689.

0006-291X/95 $5.00

the sequence elucidated by direct protein analysis of the erythrocyte enzyme has been reported. It has been proposed that G6PD in red cells is a natural "fusion" molecule, composed of 534 amino acid residues, with the first 55 residues encoded by a gene on chromosome 6 and the remaining 479 residues encoded by the classical X-linked G6PD gene (8). This report has been disputed on the basis of genetic and biochemical considerations and of experiments with recombinant protein (9) and of immunoprecipitation with specific antibodies (10). The finding has been subsequently attributed to an artifact arising during purification from frozen red blood cells after prolonged storage at -20°C (11). Therefore the actual N-terminal amino acid of the mature protein present in human erythrocytes is still undetermined.

In this paper we report the results of combined chemical and mass spectrometric analyses of human erythrocyte G6PD. These confirm the primary structure deduced from cDNA nucleotide sequence and unequivocally identify the N-terminus of the mature erythrocyte G6PD as N-acetyl-alanine.

EXPERIMENTAL PROCEDURES

Protein alkylation and digestion. The purified protein was lyophilized, redissolved in 0.1 M Tris/HCl pH 8.0, containing 6M guanidine/HCl, reduced with dithiotreitol (5-fold molar excess over thiol groups) and carboxymethylated with iodoacetic acid (20-fold molar excess over protein thiol groups). The protein was then desalted by gel filtration on a prepacked G25 column (PD-10, from Pharmacia, Uppsala, Sweden) equilibrated with 0.1% trifluoracetic acid, lyophilized and digested with trypsin (1:50 w/w) in 1% ammonium bicarbonate for 4 h at 37°C. The peptide mixture was separated by reverse-phase HPLC on a Bondapack C18 column (Waters, Milford, MA) using a System Gold apparatus (Beckman Instruments, Irvine, CA). Elution was accomplished by a linear gradient of eluant B (0.08% trifluoracetic acid in acetonitrile) in eluant A (0.1% trifluoracetic acid) at a flow rate of 1 ml/min. The eluate was monitored at 220 nm.

Amino acid analysis and sequence. Amino acid analysis was performed on an Applied Biosystems derivatizer model 420A (Applied Biosystems, Foster City, CA), equipped with an automatic hydrolysis station. Automated repetitive Edman degradation was performed on a pulsed-liquid phase sequencer model 477A from Applied Biosystems, equipped with a 120A analyzer for the on line detection of phenylthiohydantoin amino acids.

Mass spectrometry analysis. Continuous-Flow Fast Atom Bombardment (CF-FAB) spectra were performed on a VG-7070 EQ mass spectrometer (VG Analytical, Manchester, United Kingdom). Data were acquired and processed on a Vector/two system (Teknivent corp., Maryland Heights, MO). A laboratory-made CF-FAB probe, a LC-500 microflow pump (Kontron, Zürich, Switzerland), and a Rheodyne (Cotati, CA) model 7520 injector with a 0.5 l internal loop were used.
The CF-FAB matrix (acetonitrile/water 75:25 (v/v), containing 1% saturated aqueous oxalic acid and 1% 2, 2'- dithioethanol) was carried into the mass spectrometer by a deactivated fused silica capillary column (1m x 50 m internal diameter) at the flow rate

of 4 μl/min. An Ion Tech gun (Xenon beam operated at 8kev) was used, and the temperature of the probe tip was kept at 35C (scan range, m/z 90-1300; scan rate 10 s/decade; resolution 1500 (10% valley definition).

RESULTS AND DISCUSSION

G6PD was purified from erythrocytes obtained from expired blood units stored at 4°C, by means of the classical purification method based on affinity chromatography on 2', 5' ADP-Sepharose (12). The enzyme was pure as judged by SDS-PAGE and had an apparent molecular mass of 58 kDa.

Sequential Edman degradation failed to give identifiable residues, confirming that the N-terminus is blocked, as already reported for G6PDs from human erythrocytes (13), from rat liver (14) and from yeast (15). Attempts to remove the blocking group by means of mild acid treatment (incubation in 30 % trifluoroacetic acid for 3 h at 55°C) were unsuccessful. Analysis of peptides derived from tryptic digestion was subsequently undertaken.

The carboxymethylated protein was digested with trypsin and the peptides were separated by HPLC on a reverse-phase column. The elution profile is shown in Fig. 1.

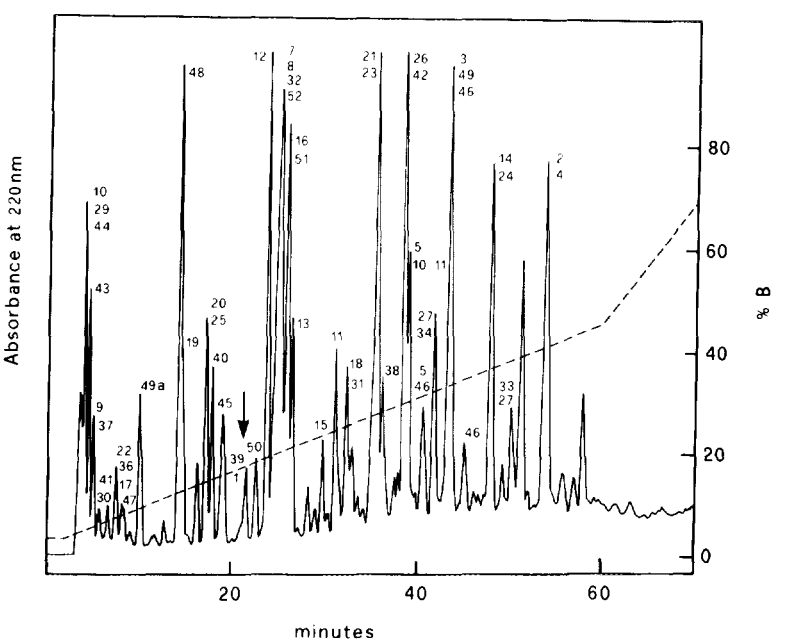

Fig. 1. HPLC elution pattern of peptides derived from trypsin digestion of carboxymethylated G6PD. The dashed line represents the elution gradient. The arrow indicates the peak which has been subjected to amino acid sequence and mass spectrometry analysis. The peptide peaks are numbered according to Fig. 2.

Identification of the peptides contained in individual peaks was obtained by amino acid analysis after acid hydrolysis and by limited Edman degradation; peptides are numbered according to their order in the sequence deduced from cDNA analysis, as shown in Fig. 2. On the basis of the amino acid composition and of the cDNA-deduced amino acid sequence more than 60 peptides were expected, ranging from 2- to 28-residues. The presence of many adjacent lysine and/or arginine residues gave rise to incomplete trypsin cleavages and therefore to peptide heterogeneity and reduced yield. The cysteine-containing peptides also were recovered in low yield, since the cysteine-alkylation reaction was not complete and peptides containing non-alkylated cysteine were lost on HPLC runs. Therefore the peptide pattern was quite complex, and

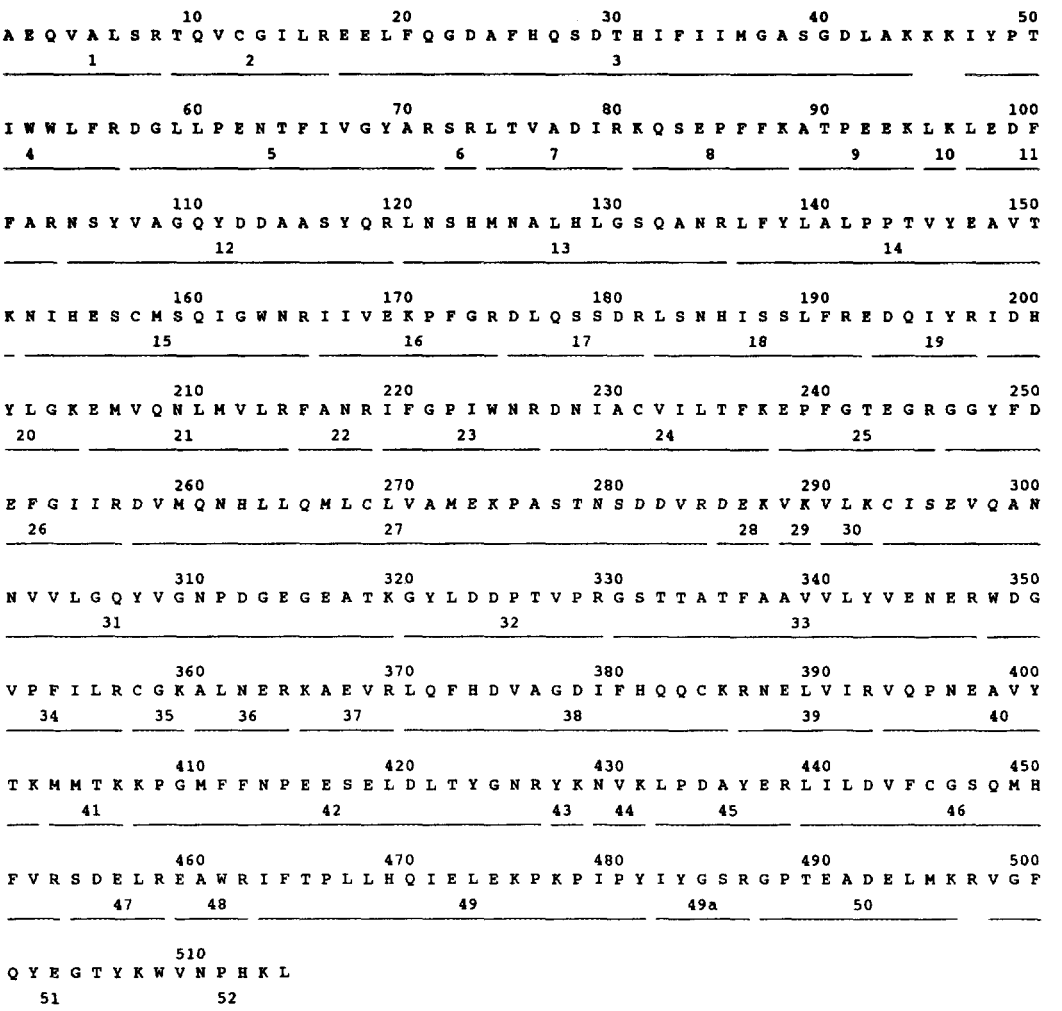

Fig. 2. G6PD amino acid sequence deduced from cDNA sequence (6). The peptides expected from trypsin digestion of the protein are underlined and numbered.

many HPLC peaks contained two or more tryptic peptides. Nonetheless, when amino acid composition was not sufficient to allow peptide identification, few cycles of Edman degradation with simultaneous amino acid sequencing allowed identification of the individual peptides even when present in mixtures.

Amino acid composition of the peak indicated by the arrow suggested the presence of two peptides in slightly different proportions, i.e. peptide 39 (150 pmoles) and peptide 1 (200 pmoles), the latter corresponding to the N-terminal peptide of the sequence deduced from cDNA, devoid of the initiator methionine (Table 1). Extensive Edman degradation produced the following single sequence: R N E L V I R, corresponding to full length peptide 39, this indicating that the other peptide had a blocked N-terminus. In order to confirm the identity of the blocked peptide, the HPLC peak was analyzed by CF-FAB-MS (Fig. 3). The analysis indicated the presence of two ions at m/z 899 and 915. The first m/z 899 ($M+H^+$) corresponds to the sequenced peptide 39; the second ion m/z 915 is 42 mass units higher than that expected for peptide 1 (Fig. 2), corresponding to the addition of an acetyl group. Therefore the N-terminal amino acid of the mature protein present in erythrocytes is N-acetyl- alanine. Accordingly, the protein has undergone post- translational cleavage of the initiator methionine and acetylation of the subsequent alanine, in accordance with predictive rules for cytosolic N-terminal processing (16). Numbering of the mature protein starts at this alanine, and the total

Table 1. Amino acid content of the peak indicated by the arrow

Amino acid	pmoles		
	total	peptide 39	peptide 1
Asp	145	145 (1)*	
Glu	480	120 (0.8)	360 (1.8)
Ser	195		195 (1)
Arg	474	300 (2)	174 (0.9)
Ala	434		434 (2.2)
Val	330	130 (0.9)	200 (1)
Ile	128	128 (0.9)	
Leu	400	150 (1)	250 (1.2)

*Numbers in parentheses indicate the moles of amino acid residue per mole of peptide.

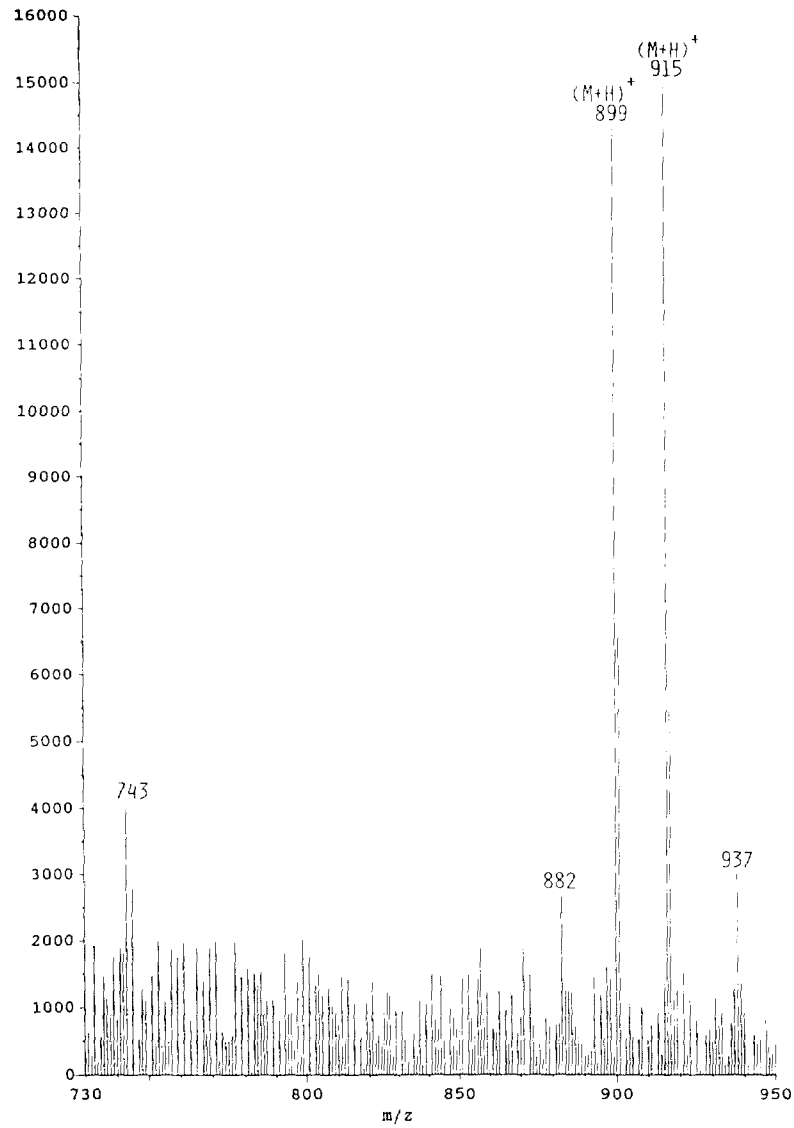

Fig. 3. FAB-MS spectrum of the HPLC peak indicated by the arrow.

amino acid sequence consists of 514 amino acid residues, including the C-terminal leucine. A probably acetylated alanine has been reported for the G6PD from rat liver (14), which has 94% residue identity with the human G6PD sequence; the N-terminal residue of the yeast G6PD is N-acetyl-serine (15).

Identification of the HPLC position of most tryptic peptides of carboxymethylated G6PD also provides a kind of "fingerprint" of normal human G6PD, which can be useful for confirmation of the structure of the mature protein. A mass profile fingerprinting

with complete identification of peptides using HPLC coupled with Electrospray ionization (ESI) or Matrix assisted laser desorption time of flight (MALDI-TOF) mass spectrometry (17) would be essential for characterization of amino acid substitutions in G6PD structural variants, where the availability of sample amount is critical. Data reported in this paper provide direct and unequivocal identification of the N-terminal amino acid sequence of human erythrocyte G6PD, which had been unidentified for many years, thus giving rise to conflicting reports.

ACKNOWLEDGMENTS Supported by Progetto Finalizzato Ingegneria Genetica (CNR). We are indebted to Dr. G. Pieraccini of Centro Interdipartimentale di Spettrometria di Massa, University of Firenze, for mass spectrometry analyses, and to Prof. A. De Flora for useful advice.

REFERENCES

1. Bonsignore, A. and De Flora, A. (1972) in Current Topics in Cellular Regulation (Horecker, B.L. and Stadtman, E.R.) 6, pp. 21-62, Academic Press, New York.
2. Luzzatto, L. and Battistuzzi, G. (1985) Adv. Hum. Genet. 14, 217-329.
3. Beutler, E. (1978) in The Metabolic Basis of Inherited Disease, 4th edn (Stanbury, J.B., Wyngaarden, J.B. and Fredrieson, D.S., eds) pp.1430-1451, Mc Graw-Hill, New York.
4. Shreve, D.S. and Levy, H.R. (1977) Biochem. Biophys. Res. Commun. 78, 1369-1375.
5. Takizawa, T., Huang, I.-Y., Ikuta, T. and Yoshida, A. (1969) Proc. Natl. Acad. Sci. USA 83, 4157-4161.
6. Persico, M. G., Viglietto, G., Martini, G., Toniolo, T., Paonessa, G., Moscatelli, C., Dono, R., Vulliamy, T., Luzzatto, L. and D'Urso, M. (1986) Nucl.Acids Res. 14, 2511-2522, 7822.
7. Descalzi-Cancedda, F., Caruso, C., Romano, M., di Prisco, G. and Camardella, L. (1984) Biochem. Biophys. Res. Commun. 118, 332-338.
8. Kanno, H., Huang, I-Y., Kan, Y. W. and Yoshida A. (1989) Cell 58, 595-606.
9. Mason, P. J., Bautista, J. M., Vulliamy, T. J., Turnar, N. and Luzzatto L. (1990) Cell 62, 9-10.
10. Beutler, E., Gelbart, T. and Kuhl, W. (1990) Cell 62, 7-9.
11. Yoshida, A. and Kan, Y.W. (1990) Cell 62, 11-12.
12. De Flora, A., Morelli, A., Benatti, U., and Giuliano, F. (1975) Arch. Biochem. Biophys. 169, 362-363.

13. Luzzatto, L., Mehta, A. (1989) in Metabolic Basis of Inherited Disease 6th edn (Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle D. eds) vol 1, pp. 2237-2265, Mc Graw-Hill, New York.
14. Jeffery, J., Barros-Söderling, J., Murray, L., Wood, I., Hansen, R. Szepesi, B. and Jörnvall, H. (1989) Eur. J. Biochem. 186, 551-556.
15. Jeffery, J., Persson, B., Wood, I., Bergman, T., Jeffery, R. and Jörnvall, H. (1993) Eur. J. Biochem. 212, 41-49.
16. Flinta, C., Persson, B., Jörnvall, H. and von Heijne, G. (1986) Eur. J. Biochem. 154, 193-196.
17. James, P., Quadroni, M., Carafoli, E. and Gonnet, G. (1993) Biochem. Biophys. Res. Commun. 195, 58-64.