Vol. 126, No. 1, 1985 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
January 16, 1985 Pages 602-609

GAWK, A NOVEL HUMAN PITUITARY POLYPEPTIDE: ISOLATION, IMMUNOCYTOCHEMICAL LOCALIZATION AND COMPLETE AMINO ACID SEQUENCE

S. Benjannet, R. Leduc, C. Lazure, N.G. Seidah, M. Marcinkiewicz and M. Chrétien

Laboratories of Biochemical and Molecular Neuroendocrinology
Clinical Research Institute of Montreal
110 Pine Avenue West
Montreal, Canada H2W 1R7

Received December 4, 1984

Summary. During the course of reverse-phase high pressure liquid chromatography (RP-HPLC) purification of a postulated big ACTH (1) from human pituitary gland extracts, a highly purified peptide bearing no resemblance to any known polypeptide was isolated. The complete sequence of this 74 amino acid polypeptide, called GAWK, has been determined. Search on a computer data bank on the possible homology to any known protein or fragment, using a mutation data matrix, failed to reveal any homology greater than 30%. An antibody produced against a synthetic fragment allowed us to detect several immunoreactive forms. The antisera also enabled us to localize the polypeptide, by immunocytochemistry, in the anterior lobe of the pituitary gland.

The pituitary gland contains many biologically active polypeptides. In human, hormones present in the anterior lobe include: prolactin, gonadotropins, thyrotropin, growth hormone and pro-opiomelanocortin (POMC)-related peptides including ACTH and β-endorphin (2-4). The posterior lobe, innervated by neurons originating in the hypothalamus, contains oxytocin, vasopressin, neurophysins, the C-terminal of propressophysin (CPP) and dynorphin (5-8). The discovery of new peptides in the pituitary gland, as well as in other organs, is of great interest and raises questions as to their source, structure, function and mechanism of action.

MATERIALS AND METHODS

Purification. Whole human pituitaries were extracted as in (9). After crude purification on a (2.5 x 30 cm) C_{18} semi-preparative HPLC column, the immunoreactive material was rechromatographed on a Waters μBondapak C_{18} (0.78 x 30 cm) equilibrated with 20% acetonitrile (containing 0.1% trifluoroacetic acid, TFA, v/v). The linear gradient was 20-40% acetonitrile (containing 0.1% TFA) in 40 min at a flow rate of 2 ml/min. Immunoreactive fractions were pooled and repurified using a linear gradient of 30-50% acetonitrile (containing 0.13% heptafluorobutyric acid, HFBA, v/v) at a flow rate of 2 ml/min.

0006-291X/85 $1.50

Sequence analyses. Amino acid analyses were carried out at 108°C for 24 hr in 5.7N HCl containing 0.1% phenol and 0.05% β-mercaptoethanol. The amino acid analyzer was an updated Beckman 120C. The primary structure was determined by automated liquid Edman degradation and performed on a Beckman 890C sequencer using a 0.3M Quadrol program. The sequencer was equipped with a Sequemat P-6 automatic converter and a Sequemat SC-510 controller. The phenylthiohydantoins (PTH) obtained were identified by RP-HPLC (10). Ophthaldehyde was used to block contaminating sequences when positions of proline residues were known as described previously (11). Double cleavages were done at those proline positions. Proteolytic cleavage was carried out with a 1% (w/w) ratio of endoproteinase Lys-C (Boehringer Mannheim) in 10 mM Tris-HCl pH 8.0 for 6 hr at 37°C. The fragments produced were purified by RP-HPLC.

Immunocytochemistry. Five µm Paraffin sections of human pituitaries preserved in Bouin's fixative were incubated for 48 hr at 4°C with rabbit antiserum and processed by the PAP method according to Sternberger (12). The optimal concentration of the antiserum diluted in PBS was 1/2000. For the control, 200 µl aliquots of diluted antiserum (1/1000 in PBS) were pre-absorbed with 4.6 µg of the synthetic antigen (residues 20-38) for 48 hr at 4°C with shaking. These were centrifuged at 12,000 rpm for 5 min and the supernatant was used for incubation. Processing with non-immunized rabbit serum and omission of the first antibody were used to determine non specific background.

RESULTS AND DISCUSSION

The complete sequence of the polypeptide which is composed of 74 amino acids is shown in Fig. 1. The polypeptide is not glycosylated, since neither glucosamine nor galactosamine were detected by amino acid analysis.

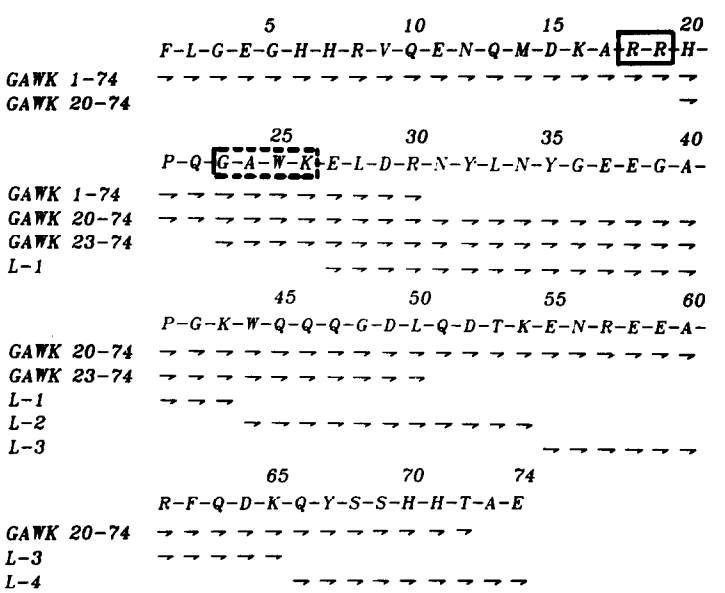

Fig. 1. The amino acid sequence of GAWK 1-74. The first peptide to be sequenced was GAWK 23-74. The second was GAWK 20-74 which added 3 more residues at the N-terminus and was sequenced up to residue 72 (see Fig 3B). A longer peptide identified as GAWK 1-74 was later isolated and the first 30 residues were identified by sequencing (Fig. 3A). The fragments denoted L-1 (3 nmol) , L-2, L-3 and L-4 were produced by enzymatic cleavage of segment 20-74 by endoproteinase Lys-C. The amino acid composition of these fragments are shown in Table 1.

The initial amino acid sequence allowed us to determine a 28 amino acid region encompassing residues 23-50 (Fig. 1). The first four amino acids of this sequence (Gly-Ala-Trp-Lys) prompted us to call this novel polypeptide GAWK (one-letter amino acid notation). Further sequence analyses permitted the chemical synthesis of a fragment corresponding to residue 20-38. Because of the very limited amount of natural material available,synthesis of this peptide was necessary, to allow the production of an antiserum and the development of a radioimmunoassay. This, in turn, helped us to purify more material as shown in Fig. 2. A single immunoreactive peak was isolated by RP-HPLC using a 0.1% trifluoroacetic acid (TFA)-acetonitrile system (Fig. 2A). Further purification using a 0.13% heptafluorobutyric acid (HFBA)-acetonitrile system led to the identification of two major immunoreactive peaks eluting at 44.4 min and 50.2 min respectively (Fig. 2B). Automatic Edman degradation carried out on the early eluting peak allowed us to confirm our previous results since that peptide starts also at position 20 and extends up to position 72 (Fig. 1). As shown in Fig. 3B, the repetitive and initial yields for that sequence were computed to be 95% and 3 nmol respectively. On the other hand, amino acid sequencing of the later eluting peak (Fig. 3A) allowed us to locate it at the NH_2-terminal part of the molecule since there was an overlap region corresponding to the sequence His-Pro-Gln-Gly-Ala-Trp... (positions 20-30, Figs. 1 and 3A). The repetitive and initial yield for this sequence were computed to be 89.1% and 10 nmol respectively. On the basis of both sequences and on the amino acid composition of the longest form available starting at Phe-Leu-Gly..., 72 amino acids were thus identified out of a possible 74. The remaining two amino acids, together with confirmation of the proposed sequence, were identified by means of a Lys-C protease digestion of the polypeptide corresponding to residues 20-74 followed by characterization of the resulting fragments. The amino acid compositions, as well as the sequence of most of these fragments, are listed in Table 1 and Fig. 1 respectively. The fragment denoted L-4 was sequenced and the results confirm the presence of two additional residues Ala-Glu following the Thr residue at position 72 (Fig. 1).

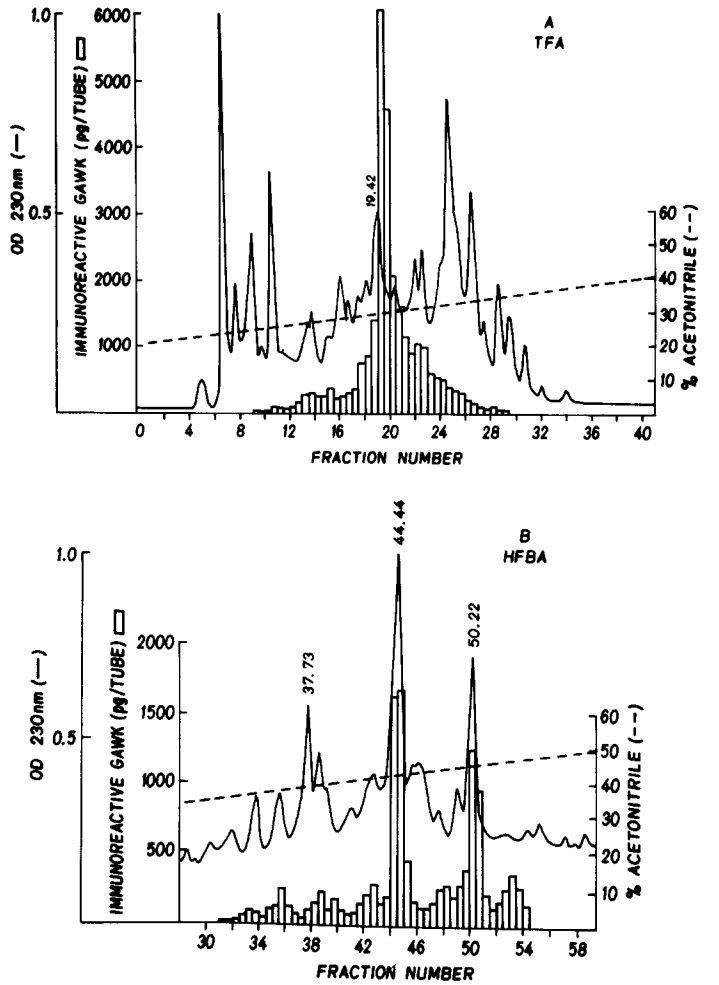

Fig. 2. Reverse-phase HPLC purification of the polypeptide from whole human pituitary extracts. Chromatogram A depicts the HPLC purification of immunoreactive material previously enriched on a 2.5-cm semi-preparative column (not shown). The column used was a Waters μBondapak C_{18} (0.78 x 30 cm) with 0.1% TFA/acetonitrile as eluent. Immunoreactive fractions (shown as histograms) were pooled and repurified using 0.13% HFBA/acetonitrile (chromatogram B). The dashed lines represent the linear gradient used in each case. The elutions were made at room temperature and monitored by the absorbance at 230 nm.

A computer data bank search (13) using a mutation data matrix developed at the National Biomedical Research Foundation (Georgetown University, Washington, D.C.) revealed no significant homologies (i.e. greater than 30%) to other known proteins or fragments thereof. Only human prosomatotropin, proalbumin and α-fetoprotein precursor resemble the region 15-22 of GAWK encompassing, interestingly, a pair of basic residues. This comparison thus

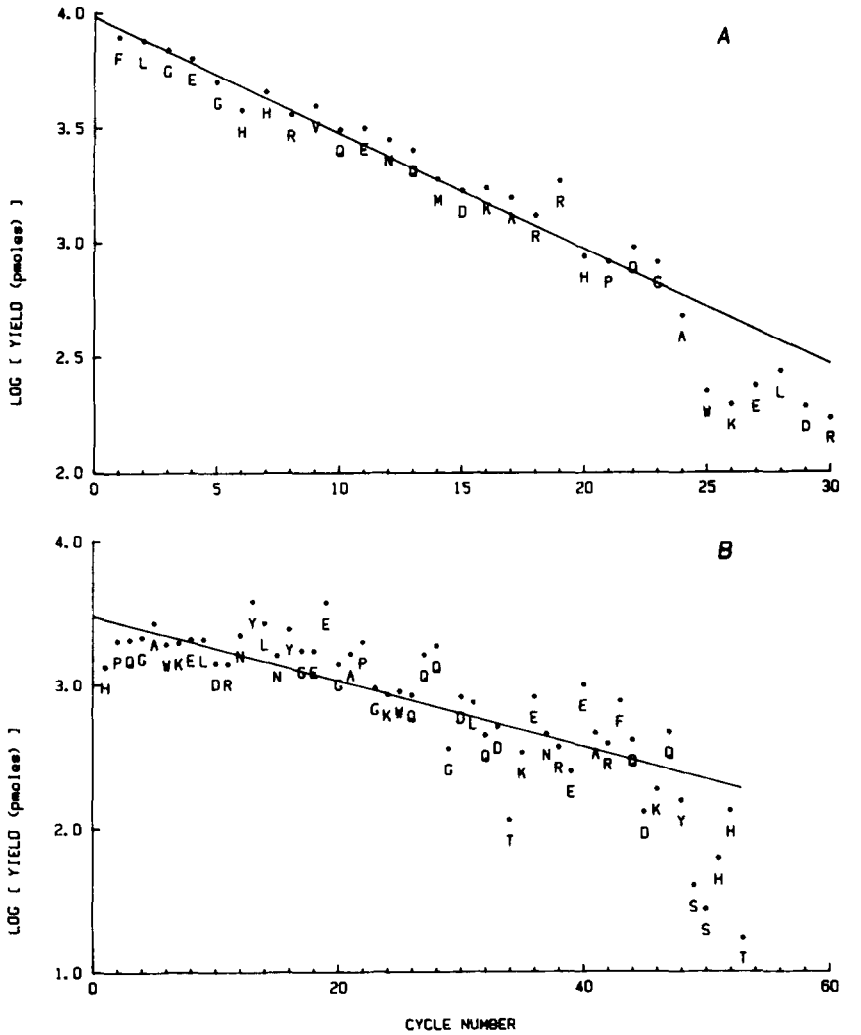

Fig. 3. The phenylthiohydantoin yields are represented as a function of sequenator cycle number obtained during the sequencing of (A) the native polypeptide and of (B) fragment 20-72. The deduced amino acid at each cycle of the sequence run is depicted in one letter notation. The solid line represents the average decline in yield along the sequence, giving an average repetitive yield of 89.1% (A) and 95% (B).

confirms the novel character of the peptide, which could represent an entirely new family of proteins.

One interesting feature resulting from the sequence of this polypeptide is the presence of numerous acidic residues (Fig. 1), Glu in particular, which represents 20% of the amino acids present. Also the degree of polarity (14), which is expressed as the ratio (in %) of Glu, Asp, Thr, Ser, Lys, His, Arg to the total number of amino acids, is high (46%). Moreover,

TABLE 1

AMINO ACID COMPOSITION OF ISOLATED ENDOPROTEINASE LYS-C FRAGMENTS AND NATIVE POLYPEPTIDE

Amino Acids	L-1	L-2	L-4	Whole Molecule	Found in Sequence
Asx	3.2 (3)	2.0 (2)	-	9.1 (9)	9
Thr	-	1.0 (1)	1.0 (1)	2.0 (2)	2
Ser	-	-	1.9 (2)	2.1 (2)	2
Glx	3.6 (3)	4.4 (4)	2.3 (2)	18.1(18)	18
Pro	1.0 (1)	-	-	2.3 (2)	2
Gly	3.4 (3)	1.3 (1)	-	7.1 (7)	7
Ala	1.2 (1)	-	1.0 (1)	4.5 (5)	5
Val	-	-	-	1.3 (1)	1
Met	-	-	-	0.9 (1)	1
Ile	-	-	-	-	-
Leu	2.2 (2)	1.0 (1)	-	4.4 (4)	4
Tyr	1.6 (2)	-	1.0 (1)	3.0 (3)	3
Phe	-	-	-	2.1 (2)	2
Lys	1.2 (1)	0.9 (1)	-	5.5 (5)	5
His	-	-	1.9 (2)	5.0 (5)	5
Arg	1.0 (1)	-	-	6.4 (6)	6
Trp	*	*	*	*	2
Total					74

* Not determined

the presence of paired residues, some of which appear twice (Fig. 1), such as His-His, Glu-Glu, Ser-Ser, Gln-Gln, is intriguing. Another important characteristic which was mentionned previously, is the presence of a pair of basic residues, Arg-Arg, at position 18-19 (Fig. 1), a feature commonly observed in hormones or biologically active peptide precursors (15). Interestingly, one of the major forms identified in this study corresponds precisely to the fragment 20-74 which is to be expected from cleavage after the pair of basic residues.

We observed positive immunoreactive material in cells on sections of the anterior lobe of the human pituitary, but very little in the posterior lobe. The observed immunoreaction in the anterior lobe showed dark-brown granules dispersed inside the cytoplasm, in contrast to the negative reaction of the nuclei. Two main areas of localization are evident: within the cells

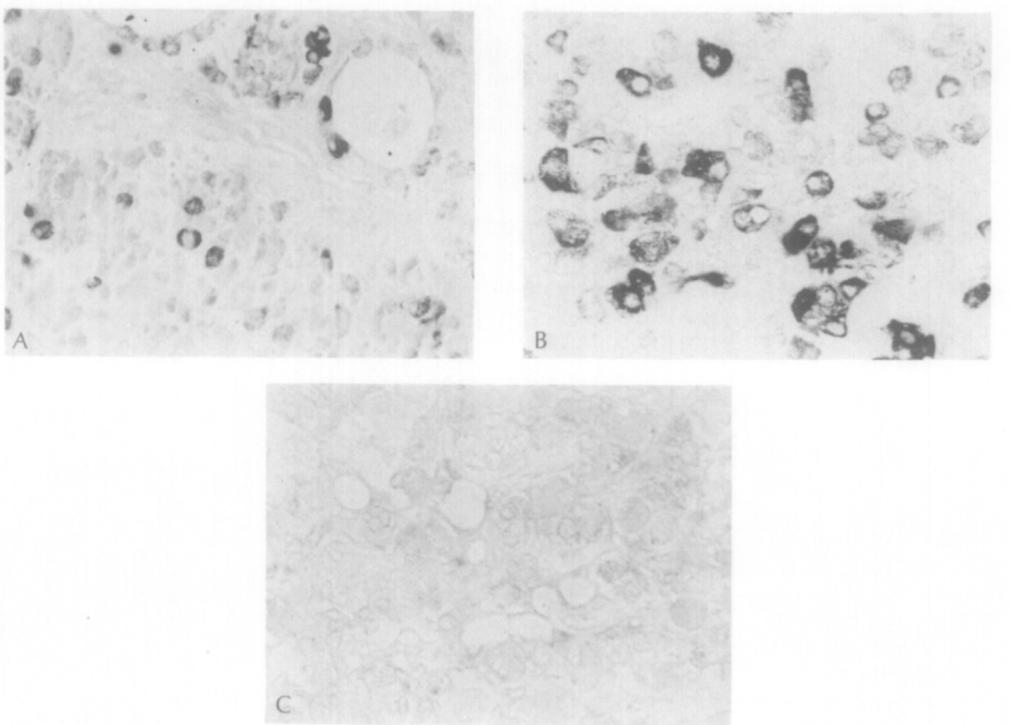

Fig. 4. Immunocytochemistry in the anterior lobe of the human pituitary gland (A, B). Dilution of the antiserum was 1/2000 in PBS. Control (C) is done by pre-incubating the antiserum (1/1000 in PBS) with synthetic antigen. A x 250. B, C x 400.

associated with the sinusoids near the posterior lobe and the stalk (Fig. 4A), and within the cells sparsely dispersed in the whole anterior part of the gland. However, the major immunoreactivity is located mainly in the ventral part of the gland (Fig. 4B). The pre-adsorption of antiserum with synthetic antigen (residues 20-38) results in a marked reduction of immunoreaction relative to non-specific background (Fig. 4C). Since the pituitary contains at least six different types of glandular cells, further immunocytochemical studies are needed to determine whether this peptide occurs in known categories of secretory cells, as was demonstrated by Nakane (16) for LH and FSH in one type of gonadotrophs.

Nucleotide back-translation of the protein sequence of GAWK reveals that the segment 44-47 exhibits the least degeneracy in the mRNA code. A chemically synthesized probe corresponding to this region will be

used to detect its mRNA in extracts of different tissues. Eventual cloning of the mRNA should allow us to answer questions regarding the structure of its precursor form, and in situ hybridization should confirm its sites of synthesis.

ACKNOWLEDGMENTS

We thank Dr. Serge St-Pierre for synthesis of peptide 20-38. We are grateful to D. Marcil and A. Sarcone for typing this report. C.L. is a "chercheur-boursier" from the FRSQ. R.L. is a FRSQ pre-doctoral fellow. This work was supported by a Program Grant (PG-2) from the Medical Research Council of Canada.

REFERENCES

1. Yalow, R.D., and Berson, S.A. (1971) Biochem. Biophys. Res. Commun. 44, 439-445.
2. Li, C.H. (1975) In: Hormonal Proteins and Peptides (C.H. Li, ed.) Vol. 3, pp. 1-40, Academic Press, New York.
3. Sairam, M.R., and Papkoff, H. (1974) In: Handbook of Physiology (E. Nobil and W.H. Sawyer, eds.) Vol. 4, p. 111, American Physiological Society, Washington, D.C.
4. Boileau, G., Seidah, N.G., and Chrétien, M. (1981) In: Hormonal Proteins and Peptides (C.H. Li, ed.) Vol. 10, pp. 65-87, Academic Press, New York.
5. Brownstein, M.J., Russell, J.T., and Gainer, H. (1980) Science 207, 373-378.
6. Seidah, N.G., Benjannet, S., and Chrétien, M. (1981) Biochem. Biophys. Res Commun. 100, 901-907.
7. Watson, S.J., Seidah, N.G., and Chrétien, M. (1982) Science 217, 853-854.
8. Watson, S.J., Akil, H., Ghazarossian, V.E., and Goldstein, A. (1981) Proc. Natl. Acad. Sci. USA 78, 1260-1263.
9. Rondeau, N., Benjannet, S., Théberge, O., Seidah, N.G., and Chrétien, M. (submitted to Anal. Biochem.)
10. Lazure, C., Seidah, N.G., Chrétien, M., Lallier, R., and St-Pierre, S. (1983) Can. J. Biochem. Cell Biol. 61, 287-292.
11. Lazure, C., Seidah, N.G., Chrétien, M., Thibault, G., Garcia, R., Cantin, M., and Genest, J. (1984) FEBS Lett. 172, 80-86.
12. Sternberger, L.A. (1979) In: Immunocytochemistry, John Wiley and Sons, New York.
13. Dayhoff, M.D. et al. (1981) Protein Sequence Data Base, Atlas of Protein, Sequence and Structure, National Biomedical Research Foundation, Washington, DC (search of June 20, 1984).
14. Nomata, Y., Watanabe, T., and Wada, H. (1983) J. Biochem. 94, 1269-1278.
15. Chrétien, M., Boileau, G., Lazure, C., and Seidah, N.G. (1984) In: Cell Biology of the Secretory Process (M. Cantin, ed.) pp. 214-246, Karger, Basel.
16. Nakane, P.K. (1970) J. Histochem. Cytochem. 18, 9-20.