Isolation and NH_2-terminal sequence of a novel porcine anterior pituitary polypeptide

Homology to proinsulin, secretin and Rous sarcoma virus transforming protein TVFV60

K.L. Hsi*, N.G. Seidah, G. De Serres and M. Chrétien+

Clinical Research Institute of Montreal, 110 Pine Avenue West, Montreal H2W 1R7, Canada

Received 27 August 1982

An M_r 21 000 polypeptide, designated APPG, has been purified by reverse-phase, high-performance liquid chromatography (RP-HPLC), from acid extracts of porcine anterior pituitary glands. This acidic protein possesses an isoelectric point of 4.9. Amino acid analysis shows that it is not a glycoprotein and estimates it to contain about 173 amino acids. NH_2-terminal sequence analysis allowed the determination of the first 50 residues unambiguously. A computer data bank search using a mutation data matrix and comparison with 269 012 protein segments indicated that this is a novel polypeptide sequence. However, this search revealed suggestive sequence homologies to a number of peptides of known sequence, including duck proinsulin (30%), Rous sarcoma virus transforming protein TVFV60 (24%) and pig secretin (26%).

Anterior pituitary peptide sequence *Proinsulin* *Secretin* *Rous sarcoma virus transforming protein*

1. INTRODUCTION

The anterior lobe of the pituitary is a source of a number of biologically active polypeptides including growth hormone, prolactin, the gonadotropins, thyrotropin and pro-opiomelanocortin (POMC), the precursor of adrenocorticotropin and β-endorphin. During the course of reverse-phase, high-performance liquid chromatography purification of the NH_2-terminal glyco-segment of POMC [1] from porcine anterior pituitary HCl/acetone extracts, a hydrophobic peptide eluting close to the porcine growth hormone position was isolated. Since we were looking for a shorter and more hydrophobic NH_2-terminal segment of POMC [2] we thought that the isolated polypeptide was related to the sought fragment. However, this turned out not to be the case, but rather a novel polypeptide was isolated. This report deals with the isolation and partial structural characterization of this protein. From the primary structure obtained a number of sequence homologies were deduced with proinsulin, secretin and Rous sarcoma virus transforming protein TVFV60.

* Visiting scholar from: Shanghai Institute of Materia Medica, Academia Sinica, China

+To whom reprint requests should be addressed

2. MATERIALS AND METHODS

2.1. *Isolation and purification of porcine APPG*

Fresh anterior pituitary glands dissected at a local slaughterhouse and immediately frozen in liquid nitrogen provided the original source of the isolated material. Following HCl/acetone extraction [1,2], the acid–acetone powder thus obtained was then dissolved in water and the solution brought to 6% (w/v) NaCl and pH adjusted to 3.0. The precipitate obtained was centrifuged, redissolved in water and the salt removed by ultrafiltration [1,2]. The desalted and lyophilized extract was then subjected to preparative reverse-phase, high-performance liquid chromatography (RP-HPLC) on a μ-Bondapak C_{18} column (0.70 × 30 cm) using a linear gradient of 2-propanol in 0.1% trifluoroacetic acid [3] (fig.1A). Repurifica-

Published by Elsevier Biomedical Press
00145793/82/0000–0000/$2.75

tion of the material denoted APPG (anterior pituitary pig) was carried out on a μ-Bondapak C_{18} column (0.39 × 30 cm) using a linear gradient of acetonitrile in 0.1 M heptafluorobutyric acid [3] (fig.1B,C). This method allowed the isolation of ~2.5 mg APPG, from ~1 g pig anterior pituitary extract.

2.2. M_r and isoelectric point estimation

The M_r of the purified APPG (fig.1C) was estimated using slab sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) as in [4] at 3 different acrylamide gradient concentrations (10–15%, 10–20% and 15–25%) (fig.2). Two-dimensional SDS–PAGE/isoelectric focusing was done as in [5].

2.3. Amino acid analysis and sequence determination

Triplicate amino acid analysis of hydrolysates of the native and performic acid oxidized APPG were carried out at 24, 48 and 72 h in 5.7 N HCl at 108°C under vacuum, using an updated Beckman 120C amino acid analyzer (table 1).

Automated liquid phase Edman degradation was performed in duplicate on 1.4 mg of the reduced and carboxymethylated APPG, using a 0.3 M Quadrol program as in [1,2,6], on a Beckman 890C sequencer equipped with a Sequemat P-6 automatic converter. Phenylthiohydantoins obtained were identified by RP-HPLC as in [6].

3. RESULTS

The initial purification of APPG is shown in fig.1A. This peptide elutes at ~40% 2-propanol on this column. Repurification of material under the peak designated APPG is shown in fig.1B,C where, under this system, APPG elutes at ~52% of acetonitrile.

The homogeneity of the preparation was checked on a slab SDS–PAGE system as shown in fig.2. It is seen that the isolated APPG is homogeneous and migrates with app. M_r 21 000. Upon two-dimensional isoelectric focusing the peptide was also found to migrate as a single band and the estimated isoelectric point was pI 4.9 (not shown).

The amino acid composition of APPG is depicted in table 1. Excluding tryptophan, it contains 173 amino acid residues. Neither glucosamine nor galactosamine was detectable showing that APPG is not glycosylated.

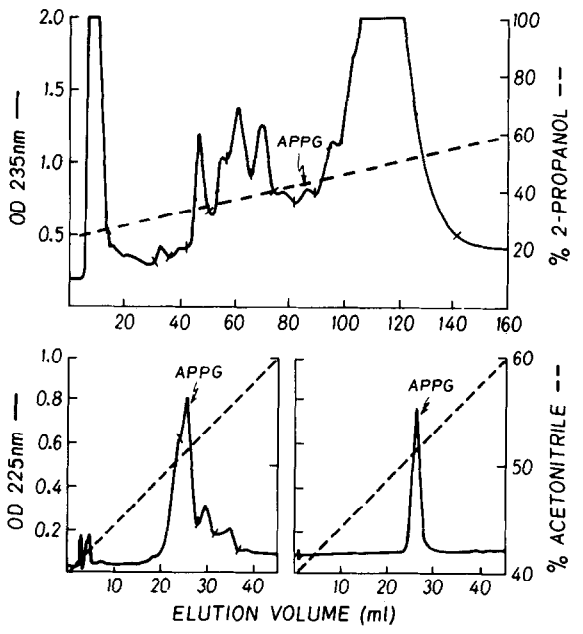

Fig.1. Reverse-phase HPLC purification of APPG from pig anterior pituitary HCl/acetone extracts using Model 322 Beckman HPLC instrument. (A) The extract was eluted from a Waters μ-Bondapak C_{18} column (0.78 × 30 cm) with a linear gradient of 2-propanol from 20–60% in 0.1% trifluoroacetic acid (TFA) at 2 ml/min. Repurification of material under the arrow is shown in (B,C). Here a Waters μ-Bondapak C_{18} column (0.39 × 30 cm) was eluted with a linear gradient of acetonitrile from 40–60% in 0.01 M heptafluorobutyric acid (HFBA) [3] at 1.5 ml/min. Material under the arrow (C) represents the APPG peptide analysed. The dashed line in each chromatogram represents the gradient used.

The PTH-amino acid yields of the NH_2-terminal sequence of APPG are shown in fig.3. The first 50 residues were identified unambiguously giving an overall repetitive yield of ~97.7%. The deduced sequence is shown in the top line of fig. 4,5.

4. DISCUSSION

In order to verify whether this is indeed a novel sequence, and if so, the homologies between this sequence and that of known proteins, a computer data bank search was performed using M.

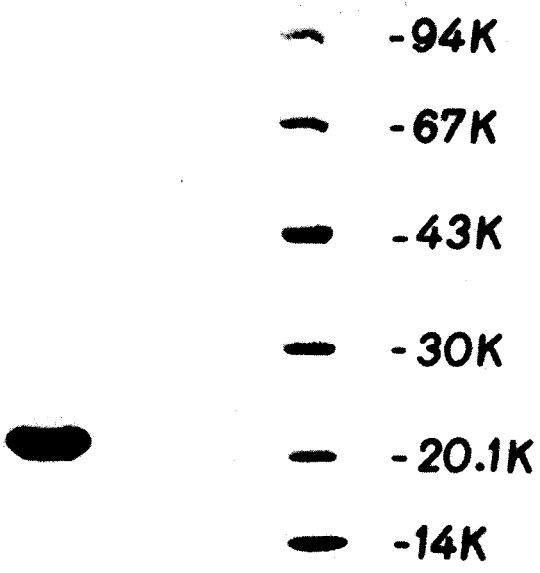

Fig.2. SDS–PAGE of APPG in a 10–20% polyacrylamide gradient slab gel. The marker proteins of known M_r used for calibration are phosphorylase *b* (94 000), bovine serum albumin (67 000), ovalbumin (43 000), carbonic anhydrase (30 000), soybean trypsin inhibitor (20 100) and α-lactalbumin (14 000).

Table 1

Amino acid composition of APPG

Amino acid	Hydrolysis time (h)			Nearest integer
	24	48	72	
Asp	21.10	20.66	20.74	21
Thr	6.92	6.49	6.40	7
Ser	7.07	6.15	5.83	7
Glu	26.54	25.86	25.66	26
Pro	16.31	16.14	17.32	17
Gly	17.95	17.89	17.62	18
Ala	9.78	9.41	9.62	10
Cys[a]	2.07	2.01	1.96	2
Val	6.55	6.78	7.15	7
Met	2.96	2.95	3.07	3
Ile	7.09	7.26	7.41	7
Leu	16.05	15.87	15.94	16
Tyr	5.53	5.39	5.52	5
Phe	5.86	5.85	5.91	6
Lys	8.49	8.35	8.40	8
His	7.72	7.56	7.56	8
Arg	7.68	7.57	7.56	8
Glu-NH_2[b]	0			
Gal-NH_2[b]	0			
Trp[c]				

[a]Value taken as oxidized derivative, cysteic acid

[b]Glu-NH_2, glucosamine; Gal-NH_2, galactosamine

[c]Not determined

Dayhoff's mutation data matrix SEARCH procedure [7,8]. The program SEARCH [8] compares a test sequence of, for example, 50 residues (as for APPG) with all 50 residue segments of the protein sequences in the data base and with progressively shorter segments at the ends of those sequences. The distribution of scores of unrelated sequences is approximately normal; related segments appear in an abnormally long tail of high scores. Typically, for a 50 residues segment, all corresponding sequences of the same protein family (sequences < 50% different) appear above the distribution of scores of unrelated segments. Therefore, the program SEARCH is used as a screening procedure to identify possible candidates for relationship to the test sequence. Unless the homology is obvious and extensive, another computer program ALIGN [8] is used to assess the probability that the sequence alignment obtained could have occurred by chance. Such a procedure measures the relatedness in standard deviation (SD) units above the distribution of chance relationships. Under such stringent search conditions, the results obtained showed that out of a total of 269 012 protein segments screened containing 1715 sequences, only proinsulin, Rous sarcoma virus transforming protein TVFV60 and secretin [7] bear suggestive sequence similarities (>20%) to this novel peptide. Furthermore, from the scores obtained using the ALIGN program for APPG comparison with duck proinsulin (2.9 SD), TVFV60 (2.6 SD) and secretin (2.3 SD), it can be argued that the probability of such similarities being due purely to chance would be $< 10^{-3}$ for duck proinsulin, $< 5 \times 10^{-2}$ for TVFV60 and $< 10^{-2}$ for secretin.

Accordingly, in fig.4 the sequence of the various proinsulins is compared with that obtained for APPG. It is seen that starting from residue 26 of the B-chain of proinsulins duck proinsulin bears the greatest homology showing 15 out of 50 residues identity. The order of homologies was: duck (IPDK 30%), chicken (IPCH 22%), rat 1 (IPRT$_1$

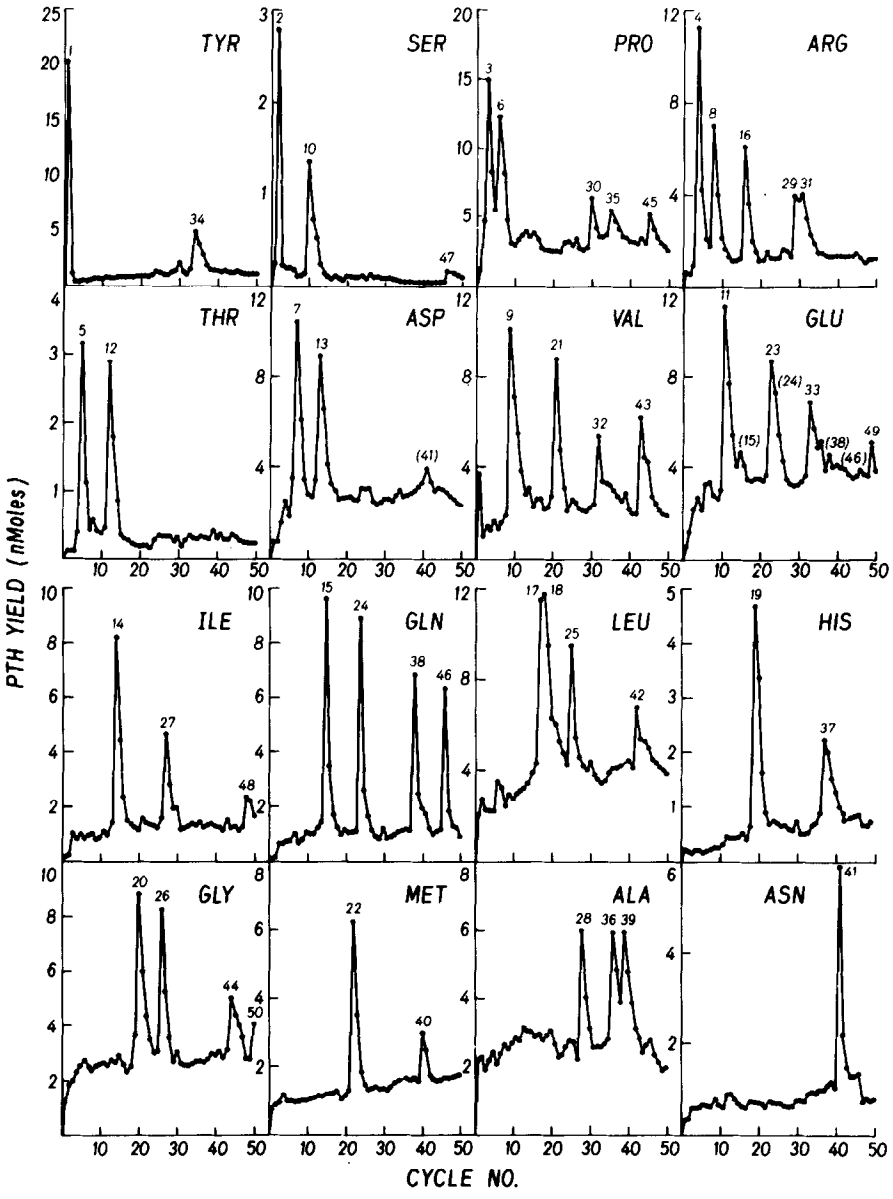

Fig.3. Yield of each PTH-amino acid obtained during sequencing of APPG, as a function of sequencer cycle number. The numbers above each peak denote the assigned sequence position of that particular residue. The numbers in parentheses for Glu and Asp mean that these residues are Gln and Asn, respectively.

20%), horse (IPHO 20%), rat 2 ($IPRT_2$ 20%), human (IPHU 18%), ox (IPBO 16%) and pig (IPPG 14%). This result also shows that the APPG peptide isolated, which contains only 2 cysteine residues, would not form the disulfide bridges found in proinsulins. Furthermore, in place of the 2 pairs of basic residues Arg–Arg and Lys–Arg found in all proinsulins joining the connecting peptide with the B and A chains (see fig.4), Pro–Asp and Met–Asn pairs are found, hence explaining the absence of cleavage of APPG at these residues [6,9].

In fig.5, comparison is made between the APPG deduced sequence and that reported for segment

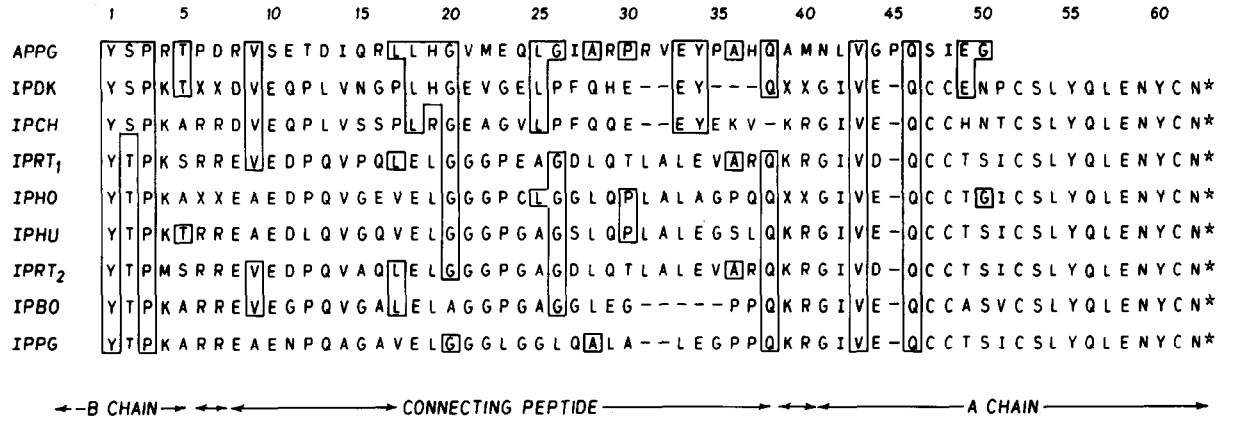

Fig.4. Deduced NH_2-terminal sequence of the first 50 residues of APPG. Best alignment of this sequence with that of proinsulins starting at residue 26 of the B-chain is also shown. Gaps have been introduced for maximum homology. Proinsulins are denoted by the first two letters (IP) and the species by the last two: duck (IPDK), chicken (IPCK), rat ($IPRT_1$ and $IPRT_2$), horse (IPHO), human (IPHU), ox (IPBO) and pig (IPPG). Identical residues are boxed in. The * represents the carboxyterminal residue.

138–190 of Rous sarcoma virus transforming protein TVFV60 [10] and whole pig secretin, SEPG [11]. It is seen that APPG contains 12 out of 50 residues identity (24%) with TVFV60 and 7 out 28 residues identity (26%) with pig secretin. That such homologies are significant is supported by the mutation data matrix SEARCH, which gave the highest match scores with duck proinsulin, TVFV60 and pig secretin [7].

Based on the properties of APPG, the partial sequence determined and the comprehensive computer data bank search results, it is clear that this is a novel sequence. The structural homology to proinsulin, TVFV60 and secretin allows us to speculate on their possible common ancestral origin and perhaps common functional properties.

In this respect, it is interesting to note that a family of polypeptides classified as transforming growth factors (TGFs) have been found in various neoplastic cells and tumors [12,13] and most recently in many non-neoplastic tissues [14,15]. These TGFs were reported to possess many similarities with APPG; all of them are acidic and heat stable and have low M_r-values (7000–23 000) and one of them, isolated from a human epidermoid carcinoma cell in [13] has the same app. M_r (21 000). Further comparison is not possible in view of the lack of sequence data on this transforming protein.

Concerning the sequence homology to proinsulins, it is noteworthy that pig proinsulin is the least homologous to pig APPG. A similar observation was reported for another anterior pituitary peptide [6], where it was found that the NH_2-terminal segment of human pro-opiomelanocortin was to some extent homologous to a number of

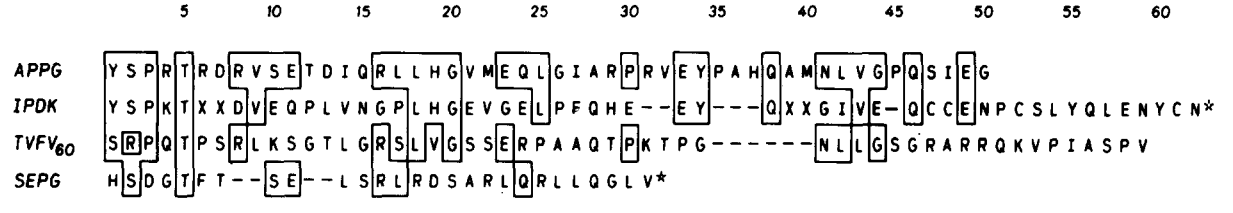

Fig.5. Alignment of the APPG sequence with duck proinsulin (IPDK) starting at residue 26; Rous sarcoma virus transforming protein (TVFV60), starting at residue 138; and pig secretin (SEPG) starting at residue 1. The * denotes the carboxyterminal residue. Identical residues are boxed in.

calcitonins but least to human calcitonin [6]. From the computer data bank search it was found that the insulin-like growth factors I and II did not show any significant homology to APPG, nor did relaxin [7]. With respect to secretin, when the complete genomic structure of pro-secretin will be known, the carboxyterminal segment could then be compared with APPG in order to ascertain the importance of such homology which is presently only with the 28 residues secretin. Therefore, it is still too early to evaluate the importance of APPG and its biological role. It is hoped that the completion of its primary structure and the screening for biological activities will shed more light on this interesting novel polypeptide.

ACKNOWLEDGEMENTS

Work supported by the Medical Research Council of Canada (PG-2), the National Cancer Institute of Canada and the National Institutes of Health (NS16315-02). We would like to thank Dr M.O. Dayhoff for performing the computer search and D. Marcil for typing the manuscript.

REFERENCES

[1] Larivière, N., Seidah, N.G. and Chrétien, M. (1981) Int. J. Pept. Prot. Res. 18, 487–491.
[2] Boileau, G., Larivière, N., Hsi, K.L., Seidah, N.G. and Chrétien, M. (1982) Biochemistry, in press.
[3] Bennett, H.B.J., Browne, C.H. and Solomon, S. (1981) Biochemistry 20, 4530–4538.
[4] Laemmli, U.K. (1970) Nature 227, 680–685.
[5] O'Farrell, P.H. (1975) J. Biol. Chem. 250, 4007–4021.
[6] Seidah, N.G., Rochemont, J., Hamelin, J., Lis, M. and Chrétien, M. (1981) J. Biol. Chem. 256, 7977–7984.
[7] Dayhoff, M.O. et al. (1981) Protein Sequence Data Base, Atlas of Protein, Sequence and Structure, version 2, Oct. 8; updated April 1982, National Biomedical Research Foundation, Washington DC; search of February 3 and April 6, 1982.
[8] Dayhoff, M.O. (1979) in: Atlas of Protein Sequence and Structure (Dayhoff, M.O. ed), vol. 5, suppl. 3, pp. 1–8, National Biomedical Research Foundation, Washington DC.
[9] Steiner, D., Quinn, P.S., Chan, S.J., Marsh, J. and Tager, H.S. (1980) Ann. NY Acad. Sci. 343, 1–16.
[10] Czernilovsky, A.P. et al. (1980) Nature 287, 198–203.
[11] Mutt, V., Jorpes, J.E. and Magnusson, S.D. (1970) Eur. J. Biochem. 15, 513–519.
[12] De Larco, J.E. and Todaro, G.J. (1978) Proc. Natl. Acad. Sci. USA 75, 4001–4005.
[13] Todaro, G.J., Fryling, C. and De Larco, J.E. (1980) Proc. Natl. Acad. Sci. USA 77, 5258–5262.
[14] Roberts, A.B., Anzano, M.A., Lamb, L.C., Smith, J.M. and Sporn, M.B. (1981) Proc. Natl. Acad. Sci. USA 78, 5339–5343.
[15] Roberts, A.B., Auzano, M.A., Lamb. L.C., Smith, J.M., Frolik, A.C., Marquardt, H., Todaro, G.J. and Sporn, M.B. (1982) Nature 295, 417–419.