Proc. Nat. Acad. Sci. USA
Vol. 68, No. 4, pp. 866–869, April 1971

Sequences of Pituitary and Placental Lactogenic and Growth Hormones: Evolution from a Primordial Peptide by Gene Reduplication

H. D. NIALL, M. L. HOGAN, R. SAUER, I. Y. ROSENBLUM, AND F. C. GREENWOOD

Endocrine Unit, Massachusetts General Hospital, Boston, Mass. 02114; and Department of Biochemistry and Biophysics, University of Hawaii, Honolulu, Hawaii 96822

Communicated by Paul C. Zamecnik, February 10, 1971

ABSTRACT **Human placental lactogen has been found to resemble human pituitary growth hormone very closely in amino acid sequence, about 80% of the residues examined being identical in the two molecules when a revised sequence for growth hormone is used as the basis for comparison. The structural features responsible for the differing biological potency of the two hormones may therefore reside in rather limited regions of primary structure. The observation of internal sequence homologies within the pituitary growth hormone and prolactin and the placental lactogen molecules suggests that these polypeptide hormones may have evolved by genetic reduplication from a smaller common ancestral peptide. This finding directs further attention to subfragments of these molecules as possible possessors of intrinsic somatotrophic and lactogenic activity.**

Previous studies on pituitary growth hormone and prolactin, and on placental lactogen, have demonstrated a close structural similarity within this group of hormones, and thus provided a basis for their shared biological and immunological properties (1–6). However, complete amino acid sequences have been reported so far only for human growth hormone (1) (growth hormone) and ovine prolactin (3) (prolactin), and detailed intra-species comparisons of structure have not been possible. Recently, we have extended earlier structural studies (2, 5) on human placental lactogen (lactogen) and have determined much of its amino acid sequence. In the course of this work it was noted that discrepancies existed between the amino acid sequence of growth hormone as previously reported by Li and coworkers (1), and that which would have been predicted by homology with the structure of lactogen as determined in our laboratory. This led us to postulate an error in the previous growth hormone sequence, and to undertake a reinvestigation of its primary structure. Our results (7) showed that the previous amino-terminal structure of growth hormone was in fact incorrect, and that a sequence of 15 amino acids containing the single tryptophan of growth hormone, assigned by Li and coworkers to positions 17–31, must reside elsewhere in the molecule, most probably occupying positions 77–91. Our more recent investigations (manuscript in preparation) of the growth-hormone sequence have confirmed that this is indeed the correct location for the tryptophan sequence, and have also demonstrated the presence of the proposed (7) "missing" dipeptide sequence (Leu-Arg) at positions 92 and 93. Work in progress strongly suggests the existence of several further errors in the previous sequence for growth hormone. One of these involves residues 130–132, which we find to be Gly-Ser-Pro rather than Pro-Ser-Gly (see Fig. 4).

The present report describes sequence studies on lactogen that demonstrate an extremely close homology with the revised growth hormone structure (7). More surprisingly, we also observed unequivocal internal sequence homology between four different regions of the lactogen structure. Examination of the revised growth hormone sequence and the prolactin sequence also revealed the presence of internal homologies that had not previously been detected.

These findings suggest a possible mode of evolution for growth hormone and the lactogenic hormones from a common ancestral peptide by a series of genetic reduplications (8).

Abbreviation: TPCK, tosyl-amido-2-phenylethyl-chloromethyl ketone.

MATERIALS AND METHODS

Highly purified samples of lactogen were kindly provided by Dr. A. Parcells of the Upjohn Co. and by Dr. H. Friesen. End-group analyses by the phenylisothiocyanate method (9) were performed on all starting materials to ascertain sample homogeneity.

Succinylation of lactogen was performed in 0.1 M sodium borate buffer, pH 9.0. A 30-fold molar excess of succinic anhydride was gradually added, with constant stirring, over a period of 1 hr, with additions of alkali as required to maintain the pH between 8.5 and 9.0. The solution was dialyzed extensively against distilled water at 4°C and lyophilized.

The tryptophanyl peptide bond in lactogen was specifically cleaved by a modification of the method of Atassi (10). Samples of native or succinylated lactogen were dissolved in 0.1 M sodium borate, pH 9.0, and a 70-fold molar excess of solid sodium metaperiodate was added. The oxidation was terminated by addition of excess ethylene glycol (0.1 ml per ml of reaction solution). This mixture was then dialyzed extensively against distilled water at 4°C, and finally lyophilized. The tryptophanyl bond was cleaved in 70% formic acid for 6 hr in the dark at room temperature. Performic acid oxidation of this material was performed by a previously described method (11).

Digestion of succinylated lactogen with trypsin (Worthington, TPCK-treated) was for 4 hr at 37°C, at an enzyme:substrate ratio of 1:100, in 0.1 M trimethylamine–acetate buffer, pH 8.0. Fragments produced by tryptic digestion of succinylated lactogen were fractionated on a Sephadex G-50 column (1.2 × 140 cm) in the same buffer. Purity of isolated peptides was established by end-group analysis and

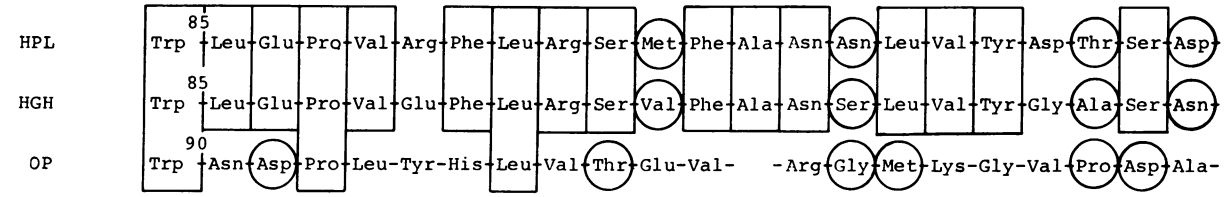

FIG. 1. Sequences of lactogen (present work), growth hormone [revised version (7, 1)], and prolactin (3) after the homologous tryptophan residue. Identical residues are enclosed in rectangles; residues related through highly favored codon substitutions (18) are enclosed in circles. A gap has been introduced in the OP sequence to improve the alignment. In this and subsequent figures, HPL denotes lactogen, HGH denotes growth hormone, and OP denotes prolactin.

amino acid compositional analysis. Fractions shown to be nonhomogeneous were further purified by ion-exchange chromatography on SE-Sephadex or DEAE-Sephadex by standard procedures using linear pH and salt gradients.

Sequence analysis

Automated Edman degradations were performed in a Beckman "Sequencer," model 890. Intact lactogen and large peptide fragments were degraded by the protein program described previously (12). Some degradations on peptide fragments were also performed with volatile reagents and procedures developed to minimize extractive losses of material (13, 14). Manual Edman degradations were performed as previously described (15). Amino acid phenylthiohydantoin derivatives were identified by gas chromatography (16) and thin-layer chromatography (12), except for the arginine and histidine derivatives, which were identified by phenanthrenequinone (17) and Pauly reactions, respectively. The phenylthiohydantoin derivative of ϵ-succinyl lysine was identified by thin-layer chromatography.

RESULTS

Analytical studies

All lactogen samples showed clear evidence of the amino terminal heterogeneity previously described (2), a sequence being present that lacked the amino-terminal Val-Gln. Quantitation of the phenylthiohydantoins by gas chromatography revealed the shortened sequence (which has amino-terminal threonine) to be present in proportions varying from 20 to 50% of the total yield at each step. No other evidence of heterogeneity was found and no contaminating proteins were detected.

Cleavage at the tryptophan residue

In preliminary experiments on native lactogen, end-group analysis of the products of cleavage at the tryptophan residue revealed the presence of only one new N-terminal amino acid (leucine). The efficiency of cleavage, measured by the ratio of amino-terminal leucine to the total of amino-terminal valine and threonine, was 40–80% in a series of five different experiments. On cleavage of succinylated lactogen, leucine was the only amino-terminal residue found in the unfractionated product; the absence of amino-terminal valine and threonine indicated that the amino groups of the starting material had been completely succinylated.

After performic acid oxidation, the two fragments (Trp-1 and Trp-2) produced by the tryptophan cleavage procedure were readily separated from uncleaved lactogen by gel filtration. However, attempts to resolve the fragments from one another by gel filtration or ion-exchange chromatography under various dissociating conditions were unsuccessful.

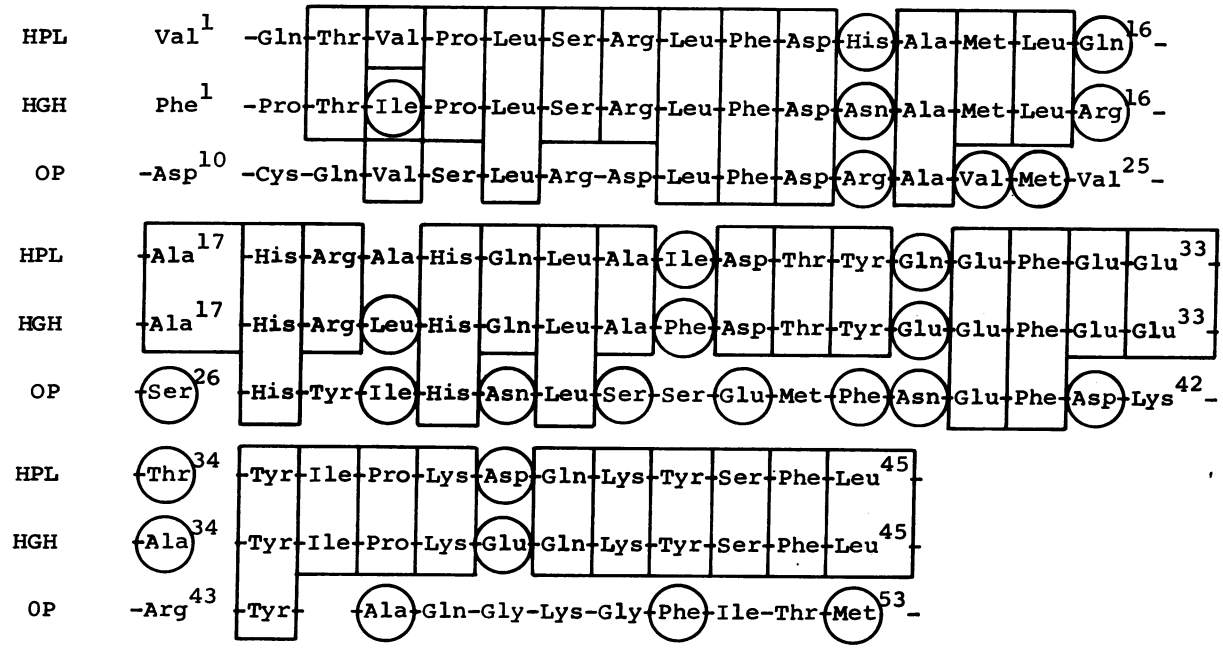

FIG. 2. Amino terminal sequences of lactogen (present work), growth hormone [revised version (7, 1)], and prolactin (3). Homologies are indicated as in Fig. 1.

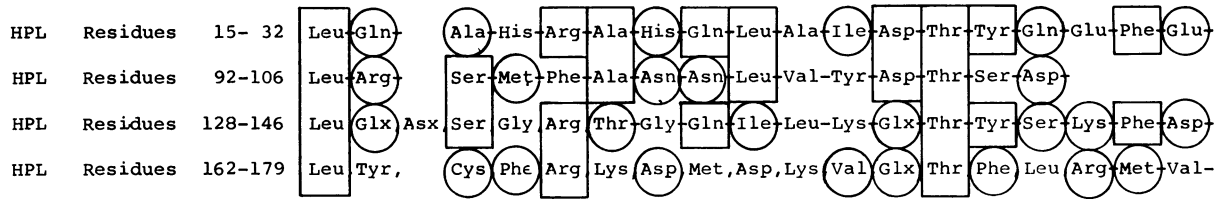

FIG. 3. Sequences of lactogen that show internal homology. Residues 15–32, 92–106, and 134–146 are from the present work. Residues 128–133 and 162–179 are assigned tentatively on the basis of our own compositional data and that of Sherwood (4). Homologies are indicated as in Fig. 1.

Subsequently, the cumulative data from amino acid composition and partial sequence of tryptic and cyanogen bromide peptides showed that the tryptophan residue occupied position 85 from the amino terminus (manuscript in preparation). This explained the difficulty in resolving Trp-1 and Trp-2, since they are very nearly equal in size and similar in charge. However, the unseparated fragments from tryptophan cleavage of succinylated lactogen were successfully subjected to automated Edman degradation for 21 cycles. Since the amino terminus of Trp-1 was blocked by succinylation, a single sequence was obtained from Trp-2; this is shown in Fig. 1, together with corresponding regions of the revised human growth hormone sequence and the ovine prolactin sequence. A close homology is evident, particularly between growth hormone and lactogen.

Sequence studies on intact lactogen and its tryptic peptides

Several automated Edman degradations were performed on native lactogen. These confirmed earlier results for the first 17 residues (2), and extended the sequence determination as far as residue 30 (Fig. 2). Sequence studies were also done on three tryptic peptides from succinylated lactogen (ST 3, ST 6, and ST 9), using the automated peptide degradation techniques. These analyses established the sequences for residues 20–45 (overlapping the amino-terminal degradations on intact lactogen; see Fig. 2) and 134–146, and confirmed the sequence for residues 94–106 (Fig. 3), which had been established through degradation of Trp-2.

DISCUSSION

The new sequence information reported here extends earlier observations of a close structural homology between human placental lactogen and human pituitary growth hormone. The sequence comparisons (Figs. 1–4) show that over 80% of the amino acids in corresponding positions are identical; amino acids pairs related through highly favored codon substitutions (18) occupy most of the remaining positions. Our preliminary results (manuscript in preparation) for the rest of the lactogen molecule indicate that this degree of similarity is seen throughout the whole polypeptide chain, and show, moreover, that growth hormone and lactogen have identically placed half-cystine residues and, most probably, identical chain length (190 amino acids). Preliminary reports on the structure of lactogen have recently been announced by two other groups. Li (19) has proposed a sequence for lactogen representing about 90% of the structure. Compositional and partial sequence data on a number of peptide fragments from lactogen have been reported by Sherwood (20). Their results agree well with ours in general, though a full comparison must await publication of more detailed reports.

Homologies between the two human hormones and ovine prolactin are also evident (Figs. 1–4), though less striking than the lactogen–growth hormone similarity. It is not possible to decide on the basis of present information whether particular differences in structure between prolactin and the other two hormones are biologically relevant, or whether they represent the accumulation of functionally unimportant mutations during evolutionary divergence of the human and ovine species. Further structural studies within a single species would help to settle this question.

The most striking finding in the present study was the observation of internal homology within the lactogen molecule. There appear to be four main regions showing homology distributed along the polypeptide chain (Fig. 3). The sequences (about 20 amino acids long) are clearly related, since almost every position contains amino acids that are identical or closely related through highly favored codon

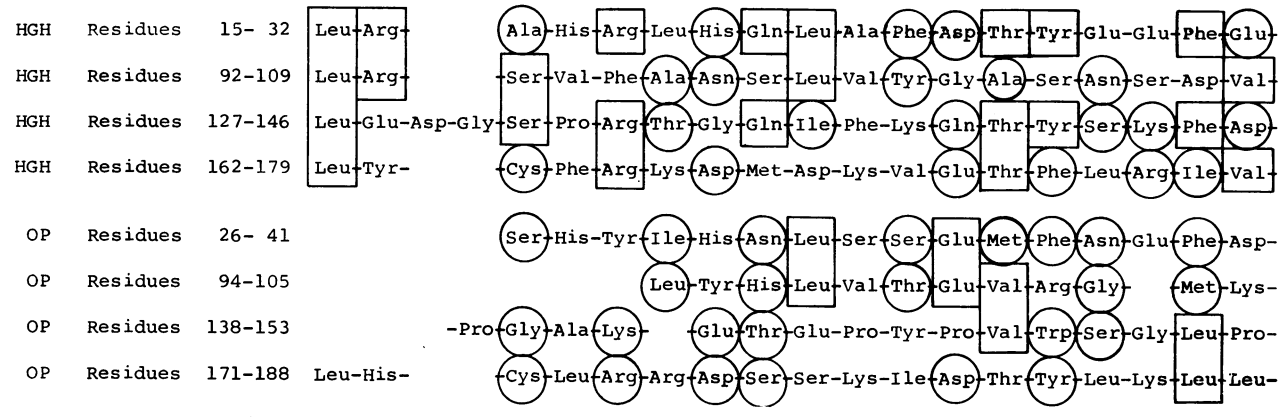

FIG. 4. Internally homologous sequences of growth hormone [revised structure (7, 1)] and prolactin (3). Homologies indicated as in Fig. 1.

substitutions. Inspection of the growth hormone and prolactin sequences shows the same regions of internal homology (Fig. 4) in corresponding positions in these molecules. This is illustrated diagramatically in Fig. 5. *External* homology between the three hormones, as opposed to the zones of *internal* homology within each hormone, is seen throughout their length.

These findings seem consistent with the hypothesis that this group of biologically related hormones, i.e., the pituitary growth hormones, the pituitary prolactins, and the placental lactogens, have arisen from a shorter primordial peptide, of perhaps 25–50 amino acids, through two successive tandem duplications of the original structural cistron. The resulting primitive polypeptide probably most closely resembled prolactin, since this hormone is found throughout the vertebrates, whereas growth hormone and placental lactogen evolved at a later stage in the higher vertebrates and mammals. Precedents exist for this line of reasoning, since there is now considerable evidence that gene reduplication was a common means of genetic diversification during evolution in the vertebrate line (21). The analogy between the present findings and those described earlier for immunoglobulin (22), hemoglobin (23), and haptoglobin (24) evolution is obvious.

In considering the significance of these findings, we think it at least plausible that a sequence corresponding to the ancestral peptide might possess demonstrable biological activity, since it must have survived the forces of natural selection long enough for duplication and reduplication to occur. Of interest is the recent report by Li and Yamashiro (25) of the synthesis of a polypeptide, based on their earlier growth hormone sequence determination, possessing both somatotrophic and lactogenic activity. As we have demonstrated (7) that this amino acid sequence is incorrect in a major respect, it would seem surprising that their product possessed any biological activity at all. However, the results could be explained by the existence of localized biologically active regions of the molecule, not disrupted by the translocation of the tryptophan-containing sequence. Other interpretations, however, are possible, and the synthesis of the correct sequence for growth hormone should prove informative. In the search for biologically active fragments, the synthesis of the internally homologous regions of growth hormone, lactogen, and prolactin is an obvious extrapolation of the present work. Bornstein and coworkers (26) have recently described the structure and synthesis of two biologically active peptides of low molecular weight from human growth hormone. One of these peptides (Ac-G) is reported to consist of residues 1–21 of the growth hormone sequence suggested by Li. Since our revision of the growth hormone structure shows that this amino acid sequence does not exist in the native molecule, its isolation is rather surprising. The report by Bornstein that both natural and synthetic versions of this peptide possess *in vitro* biological activity (26) would therefore seem to require further confirmation.

We thank Dr. R. E. Fellows, Jr. for helpful discussions. This work was supported in part by a contract grant (N.I.H. 69-2190) to one of us (FCG) from the Center for Population Research, National Institute of Child Health and Development.

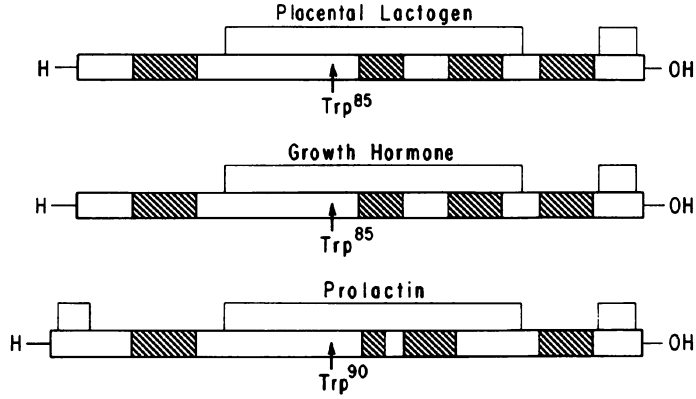

FIG. 5. Diagrammatic representation of the lactogen, growth hormone, and prolactin molecules. The cross-hatched regions of each polypeptide chain represent the internally homologous sequences shown in Figs. 3 and 4. Apart from the internal homologies within each molecule, the three hormones show considerable homology to one another, as indicated here by the similar location of disulfide bridges (*narrow lines*) and the tryptophan residues at positions 85 in lactogen and growth hormone and at position 90 in prolactin.

1. Li, C. H., J. S. Dixon, and W. K. Liu, *Arch. Biochem. Biophys.*, **133**, 70 (1969).
2. Catt, K. J., B. Moffat, and H. D. Niall, *Science*, **157**, 3786 (1967).
3. Li, C. H., J. S. Dixon, T. B. Lo, Y. A. Pankov, and K. D. Schmidt, *Nature*, **224**, 695 (1969).
4. Sherwood, L. M., *Proc. Nat. Acad. Sci. USA*, **58**, 2307 (1967).
5. Sherwood, L. M., in *Progress in Endocrinology*, ed. C. Gaul and F. J. G. Ebling (Excerpta Medical Foundation, Amsterdam, 1969).
6. Bewley, T. A., and C. H. Li, *Science*, **168**, 1361 (1970).
7. Niall, H. D., submitted to *Nature*.
8. A preliminary report of our findings concerning the lactogen sequence, the revised growth hormone sequence, and the presence of internal homologies in growth hormone, prolactin, and lactogen was presented at the N.I.H. Workshop on Prolactin, Bethesda, Md., January 11, 1971.
9. Edman, P., *Acta. Chem. Scand.*, **4**, 283 (1950).
10. Atassi, M. Z., *Arch. Biochem. Biophys.*, **120**, 56 (1967).
11. Hirs, C. H. W., *J. Biol. Chem.*, **219**, 611 (1956).
12. Edman, P., and G. Begg, *Europ. J. Biochem.*, **1**, 80 (1967).
13. Niall, H. D., H. Penhasi, P. Gilbert, R. C. Myers, F. G. Williams, and J. T. Potts, Jr., *Fed. Proc.*, **28**, 661 (1969).
14. Sauer, R., H. D. Niall, and J. T. Potts, Jr., *Fed. Proc.*, **29**, 728 (1970).
15. Niall, H. D., H. T. Keutmann, D. H. Copp, and J. T. Potts, Jr., *Proc. Nat. Acad. Sci. USA*, **64**, 771 (1969).
16. Pisano, J. J., and T. J. Bronzert, *J. Biol. Chem.*, **244**, 5597 (1969).
17. Yamada, S., and H. A. Itano, *Biochim. Biophys. Acta*, **130**, 538 (1966).
18. Dayhoff, M. O., and R. V. Eck, in *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Silver Spring, Md., 1969).
19. Li, C. H., presented at N.I.H. Workshop on Prolactin, Bethesda, Md., January 11, 1971.
20. Sherwood, L. M., presented at N.I.H. Workshop on Prolactin, Bethesda, Md., January 11, 1971.
21. Ohno, S., in *Evolution by Gene Duplication* (Springer-Verlag, N.Y., 1970).
22. Hill, R. L., R. Delaney, R. E. Fellows, and H. E. Lebovitz, *Proc. Nat. Acad. Sci. USA*, **56**, 1762 (1966).
23. Hill, R. L., J. Buettner-Janusch, and V. Buettner-Janusch, *Proc. Nat. Acad. Sci. USA*, **50**, 885 (1963).
24. Black, J. A., and G. H. Dixon, *Nature*, **218**, 736 (1968).
25. Li, C. H., and D. Yamashiro, *J. Amer. Chem. Soc.*, **92**, 7608 (1970).
26. Bornstein, J., J. McD. Armstrong, F. Ng, B. M. Paddle, and L. Misconi, *Biochem. Biophys. Res. Commun.*, **42**, 252 (1971).