Proc. Nat. Acad. Sci. USA
Vol. 69, No. 2, pp. 431–434, February 1972

Evolution of Neurophysin Proteins: The Partial Sequence of Bovine Neurophysin-I

(vasopressin/oxytocin/carrier proteins/automated amino-acid-sequence analysis/homology/protein evolution)

J. D. CAPRA*, J. M. KEHOE*, D. KOTELCHUCK†, R. WALTER†, AND E. BRESLOW‡

*Departments of Microbiology and †Physiology, The Mount Sinai School of Medicine of the City University of New York, New York, N.Y. 10029; and ‡The Department of Biochemistry, Cornell University Medical College, New York, N.Y. 10021

Communicated by Alton Meister, November 19, 1971

ABSTRACT **The sequence of the first 50 amino-acid residues of bovine neurophysin-I was determined. A comparison of this sequence with that of the 97-residue bovine neurophysin-II and the 92-residue porcine neurophysin-I molecules reveals a high degree of homology among these proteins. It is suggested that the binding site of neurophysin proteins for neurohypophyseal hormones is located in the middle portion of these molecules, where their sequences are virtually identical. The sequence data, as well as the occurrence of at least two neurophysins in both the pig and the cow, suggest that each species inherited at least two structural genes controlling the synthesis of these proteins. The most striking finding in the study was the observation of internal sequence homologies within the neurophysins. This result implies that these molecules arose by way of a series of partial gene duplications of a primitive gene that coded for a smaller ancestral protein.**

Van Dyke *et al.* (1) isolated from bovine posterior pituitary glands the neurohypophyseal hormones, vasopressin and oxytocin, as noncovalent complexes with certain protein molecules. The protein components were later given the generic name neurophysins (2), and they are considered to function as carrier substances to which the neurohypophyseal hormones are bound during transport and storage within the hypothalamo–neurohypophyseal system (3). The amino-acid sequence of one of the two major neurophysin proteins in the cow has been determined (4, 5), and the sequence of a porcine neurophysin has also been reported (6). These studies established a molecular weight of about 10,000 for the monomeric unit of neurophysins, although aggregation to higher molecular weight forms has been reported (7–9).

Native bovine neurophysin-I (BNP-I) and neurophysin-II (BNP-II) both possess one hormone-binding site per monomeric unit (10). Both neurophysins are capable of binding both oxytocin and vasopressin (7, 10). Equilibrium dialysis studies have demonstrated that the affinity constants of BNP-I or BNP-II for either oxytocin or lysine-vasopressin are essentially identical (10). This result is of particular interest, since evidence has been presented suggesting that within certain secretory neurons of the hypothalamo–neurohypophyseal system of the cow neurophysin-I is bound in vesicles with oxytocin and neurophysin-II with vasopressin (11).

In the present study the partial amino-acid sequence of BNP-I is compared with the complete amino-acid sequence of BNP-II, in an attempt to define more specifically the association between neurophysins and neurohypophyseal hormones. In addition, when the sequences of BNP-I and BNP-II are compared with the sequence of porcine neurophysin-I (PNP-I), certain aspects of the molecular evolution of the neurophysins become apparent.

Abbreviations: BNP, bovine neurophysin; PNP, porcine neurophysin; PTH, phenylthiohydantoin.

MATERIALS AND METHODS

BNP-I, purified as described by Breslow *et al.* (8), was reduced with 2-mercaptoethanol in 8.0 M urea–0.5 M Tris buffer (pH 8.2). The sulfhydryl groups were alkylated with [^{14}C]iodoacetamide (2.7 mCi/mmol) by method of Fleishman *et al.* (12). The radioactive material was desalted by gel filtration on a Sephadex G-25 column and lyophilized to yield [^{14}C]-neurophysin-I. A Beckman model 890 Protein Sequencer was used for automated Edman degradations (13, 14). The sequencer cup was loaded with 8 mg of [^{14}C]BNP-I in 0.5 ml of 1 M acetic acid. The program provided sequential additions of phenylisothiocyanate and Quadrol buffer (Wyandotte Chemical Co.), washing with benzene and ethyl acetate, double cleavage with heptafluorobutyric acid, and extraction of the thiazolinone derivative with butyl chloride. Aldehyde-free glacial acetic acid (0.1%) was added to the ethyl acetate (13) and 0.1 mM dithioerythrytol to the butyl chloride extraction solvent (15). The yield per repetitive cycle was above 95% in most instances. The yield for individual amino acids varied from 87 to 30%. To identify the amino-acid phenylthiohydantoins (PTH), a modification of the Pisano and

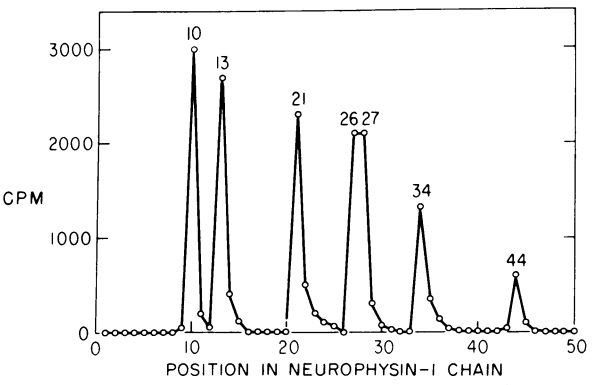

FIG. 1. Location of half-cystines in bovine neurophysin-I. A 5% aliquot of PTH-amino acid obtained from successive Edman cleavages was counted in a liquid scintillation counter.

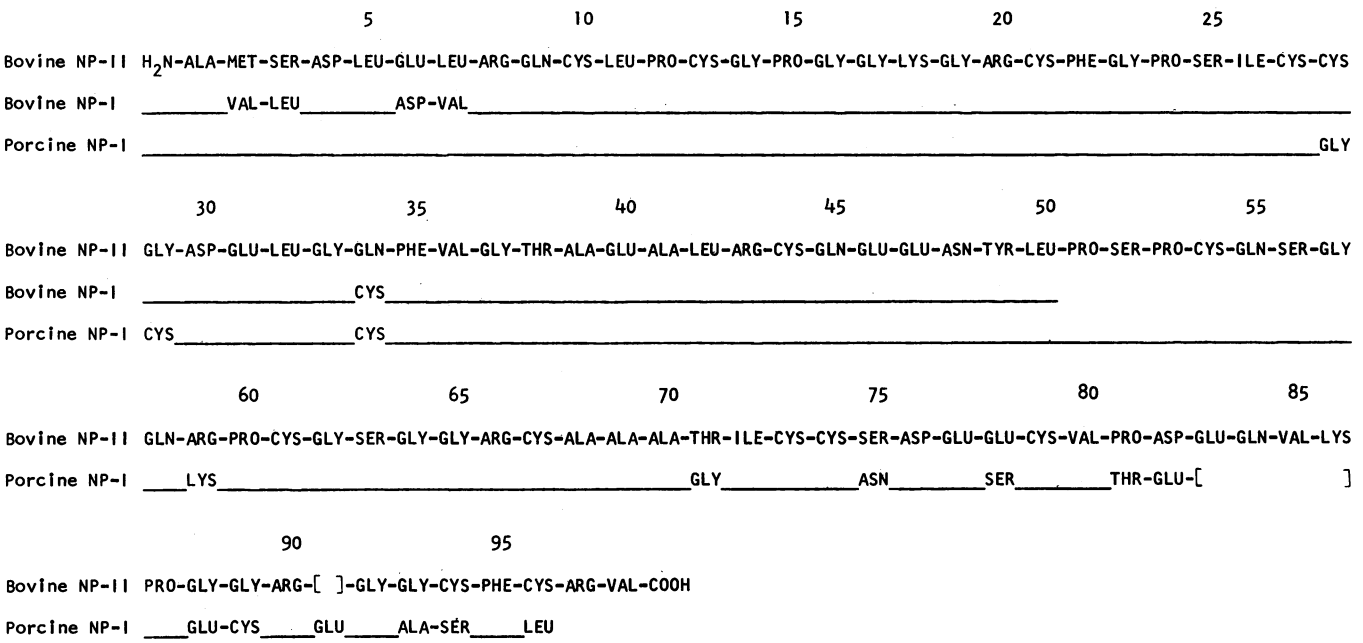

FIG. 2. Amino-acid sequences of bovine and porcine neurophysins. The BNP-II sequence is given. Only those residues in BNP-I and PNP-I that differ from BNP-II are shown. NP, neurophysin.

Bronzert (16) technique with a Beckman GC-45 gas chromatograph was used.

For certain residues, the PTH derivatives were acid hydrolysed (17), and the free amino acids were identified on a Beckman model 121 amino-acid analyzer. Other PTH derivatives were identified by a modification of the thin-layer chromatography (TLC) system of Edman and Sjoquist (18). Generally, 6% of the phenylthiohydantoin derivative was subjected to gas chromatography, 45% to TLC, and 5% to scintillation counting to facilitate the identification of half-cystine residues. When necessary, the remaining 44% was hydrolyzed and subjected to amino-acid analysis.

RESULTS AND DISCUSSION

The amino-acid sequence of BNP-I, the binding protein of oxytocin from the cow, has been determined through residue 50 by automated Edman degradation. After reduction of the disulfide bonds of the native protein with 2-mercaptoethanol, the half-cystine residues were converted to [^{14}C]labeled *S*-carboxamidomethylcysteine (H_2NCOCH_2-Cys). Such chemical modification of a native protein before sequence analysis has two advantages, as illustrated in Fig. 1. First, the cpm of ^{14}C recovered, plotted as a function of degradative cycles, indicates the efficiency of the stepwise degradation of the protein between two given half-cystine residues. This feature is particularly valuable with proteins like the neurophysins that have a high half-cystine content. In addition, the ^{14}C label provides a simple and reliable method for the detection of half-cystine residues. On gas chromatography, PTH-Ser and PTH-*S* H_2NCOCH_2-Cys are not readily distinguishable. Although TLC identifies PTH-Ser with ease, a positive identification of PTH-*S*-H_2NCOCH_2-Cys is more difficult. However, the [^{14}C]carboxamidomethylation procedure showed unequivocally that, within the first 50 amino-acid residues of BNP-I, positions 10, 13, 21, 27, 28, 34, and 44 are occupied by half-cystine residues.

An extraordinary degree of homology was evident when the partial amino-acid sequence of BNP-I was compared with that of BNP-II (Fig. 2)§. The two bovine neurophysins differ in only five of the first 50 residues (90% homology). The first seven residues of the N-terminal sequences of the two bovine proteins show four of the five changes. These changes involve positions 2, 3, 6, and 7, and can all be accounted for by single-base changes. The fifth change, from Gln to Cys at residue 34, requires a three-base mutation. These data provide strong evidence that both bovine neurophysins arose from a common precursor gene. It is possible that BNP-I and BNP-II are the products of alleles of a single genetic locus. Cases of multiple amino-acid differences among allelic genes have been reported for the hemoglobin β chains of sheep (19, 20), cattle (21, 22), and goats, as well as for bovine carboxypeptidase A (23, 24). However, the fact that BNP-I and BNP-II are found in different neurosecretory vesicles (11) makes their allelic nature unlikely.

In comparing BNP-II and PNP-I we have inserted a gap in each protein in their C-terminal regions (Fig. 2). When aligned in this manner, BNP-II and PNP-I differ in only 13 positions (87% homology)¶. The majority of these differences are in the C-terminal quarter of the molecule, since only two residues among the first 70 (97% homology) are different, as compared with eleven differences in the C-terminal 27 resi-

§ Originally (4), we reported an Asn in position 30 of BNP-II. Further studies indicate that it is an Asp residue (manuscript in preparation).

¶ Positions 27-29 show Cys-Cys-Gly in both bovine neurophysins, but a Cys-Gly-Cys sequence was reported for PNP-I (6). The validity of our result for BNP-I is confirmed by the radioactive labeling experiment shown in Fig. 1. Similar results were obtained in this region with BNP-II (4). It is conceivable that the porcine molecule has this unusual inversion, but for the present discussion we have assumed that its amino-acid sequence is also Cys-Cys-Gly.

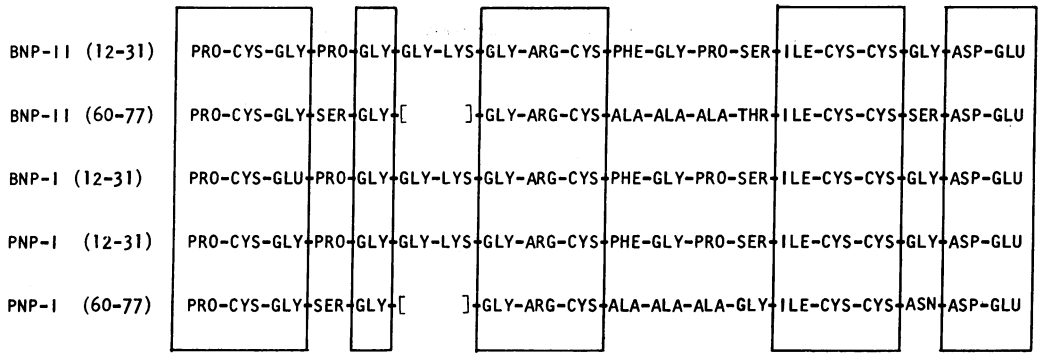

FIG. 3. Evidence for repeating sequence pattern in bovine neurophysin-II residues 12–31 and 60–77 have been aligned, and the identities are outlined.

dues (60% homology). Of the 13 differences between BNP-II and PNP-I, seven can be accounted for by single-base changes, four by two-base changes, and one (position 34) by a three-base change in the genetic code. It seems particularly noteworthy that between residues 8 and 50 all three neurophysin proteins have the same sequence, with the significant exception of residue 34. The Gln in position 34 of BNP-II was confirmed during the present study.

When all the sequence data available for bovine and porcine neurophysins are considered, the N- and C-terminal portions are seen to comprise relatively variable regions, while the central portion is essentially constant. In view of earlier findings that oxytocin and lysine-vasopressin compete for binding to native BNP-I and BNP-II (presumably for the same binding site) and show the same binding constants (10), the binding area is probably composed of constant portions of the neurophysin molecule, such as the central region.

The Cys-Gln interchange at position 34 is particularly notable. Since all the half-cystine residues in these proteins are known to be part of disulfide bonds, this interchange necessitates a difference among these proteins in the location of at least one disulfide bridge; consistent with such a change in structure, differences have been observed in the near-ultraviolet circular dichroic spectra among these proteins (8).

The sequence of BNP-II between residues 12–31 and 60–77 shows evidence of a partial gene duplication (Fig. 3). In these two stretches of about 20 residues, there are 12 identities. PNP-I shows this same duplication pattern. One might postulate that a primitive gene, which encoded a smaller 55- to 60-residue neurophysin molecule, underwent an unequal cross-over event that resulted in a partial gene duplication with fusion (Fig. 4). These findings suggest the possibility that in primitive species neurophysin molecules of less than 10,000 molecular weight might be found. A similar cross-over event has apparently occurred in human haptoglobin genes (25).

Examination of the sequence of BNP-II between residues 87–95 also suggests homology with a portion of the presumed reduplicated segment (Fig. 5). In this short stretch of eight amino acids, there are three residues that are identical in all three homology regions and six of eight residues are identical between the first and third homology regions. Of the differences, all but the Phe–Ala interchange in the last residue can be accounted for by one-base changes in the genetic code. Thus, as shown in Fig. 6, there are three homology regions in the BNP-II molecule. The first two are duplicated segments of 20 residues, and the third is a segment from the middle portion of these areas. These findings may be explained either by a partial gene triplication or by some translocational mechanism. The sequence of PNP-I in the 87–95 region shows only a very weak homology with 15–22 and 63–68, suggesting that the second cross-over may have occurred only in BNP-II, or that deletions have occurred that obscure any homology in this region in PNP-I. Elucidation of the complete sequences of BNP-I and PNP-II should be useful to further delineate the genetic mechanism involved. Niall *et al.* (26) have recently documented internal homologies in placental lactogen, growth hormone, and prolactin, revealing four homologous areas in these proteins. It is noteworthy that the first homology area in human growth hormone is between residues 15–32, similar in location to the first homology region in BNP-II [12–31]. It is also of interest that the C-terminal half of all of these proteins contains more homology areas (2 or 3) than the single area in the N-terminal half. These findings raise the possibility that a common genetic mechanism was involved in the evolution of all four of these proteins.

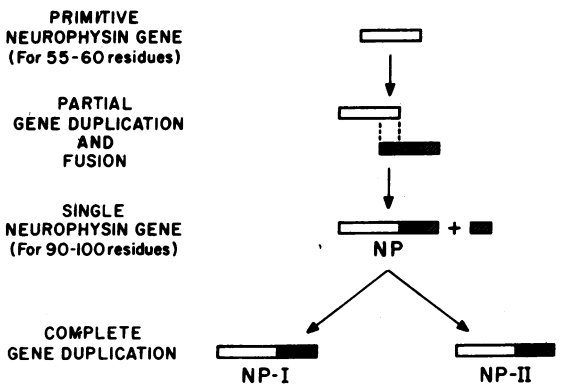

FIG. 4. Proposed evolutionary scheme for the neurophysins.

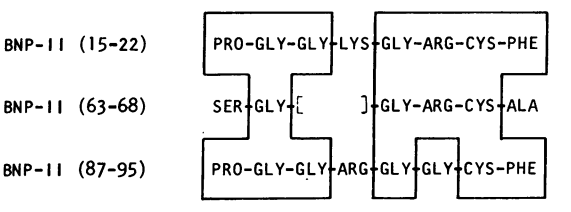

FIG. 5. Internal sequence homology of BNP-II, illustrating the three homology areas.

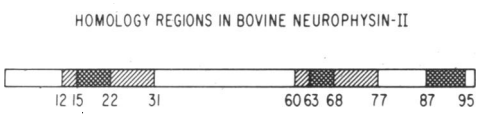

FIG. 6. Diagramatic representation of the bovine neurophysin-II molecule. *Cross-hatched* regions of the molecule represent the internally homologous sequences.

There are certain areas in the neurophysin molecules that bear a striking relationship to the neurohypophyseal hormones themselves. In one instance, it is a stretch of three amino acids of reversed polarity. Thus, residues 11–13 are Leu-Pro-Cys in all three neurophysins sequenced, while residues 6–8 of oxytocin are Cys-Pro-Leu. Moreover, in BNP-II, residues 59–61 are *Arg*-Pro-Cys; the amino-acid sequence of arginine-vasopressin (the vasopressin found in the cow) in residues 6–8 is Cys-Pro-*Arg*. In PNP-I, residues 59–61 are *Lys*-Pro-Cys. Porcine vasopressin differs from the vasopressin of most other mammals in being Cys-Pro-*Lys* in residues 6–8. The sequence of BNP-II between residues 6–14 and 54–62 shows weak homology to the sequence of arginine-vasopressin (Fig. 7). In both areas, there are three identities out of nine residues. The location of these areas of homology is precisely at the beginning of each of the areas of internal homology, and presumably the region of gene duplication. These relationships suggest the possibility that the structural genes encoding the neurohypophyseal hormones may have been incorporated into the genes that eventually gave rise to their carriers, the neurophysins.

When sets of proteins of similar function are found within the same species, they may have arisen by a process of complete gene duplication or may have evolved independently. The high degree of homology evident among the neurophysins is strong evidence for the gene-duplication hypothesis. A second question concerns whether these genes duplicated before the species diverged [as is presumably the case among the neurohypophyseal hormones (27, 28)], or whether they arose within the species, after species divergence [as is thought to have occurred with the immunoglobulin subclasses (29)]. In the cow and the pig, there are at least two neurophysins. Since among the first 50 residues there is a significantly higher degree of homology between BNP-II and PNP-I (98%) than between the two bovine neurophysins (90%), it seems somewhat more probable that neurophysin gene-duplication preceded the species divergence. The sequence of PNP-II and the complete sequence of BNP-I should allow us to choose more precisely between these two alternatives. In addition, these studies should allow a more definitive evaluation of the evolutionary significance of apparent amino-acid sequence relationships between the neurohypophyseal hormones and the neurophysins.

BNP-II (6-14)	Glu	Leu	Arg	Gln	Cys	Leu	Pro	Cys	Gly
Arginine Vasopressin	Cys	Tyr	Phe	Gln	Asn	Cys	Pro	Arg	Gly
BNP-II (54-62)	Cys	Gln	Ser	Gly	Gln	Arg	Pro	Cys	Gly

FIG. 7. The sequence of BNP-II between residues 6–14 and 54–62, showing homology to arginine-vasopressin. Sequence identities are outlined.

We thank Bonnie Gerber and Jane Weis for excellent technical assistance, Dr. Jack Ohms of Beckman Instruments, Palo Alto, Calif. for continued advice and interest, and Drs. Norman Arnheim, Robert Hill, and Henry Kunkel for helpful discussions. This work was supported by NSF grant GB-17046, USPHS grant AI-09810, and a. N.Y. Heart Association Grant in Aid to J.D.C. and J.M.K., by USPHS grant AM-13567 and Grant 1105 from The City University of New York Faculty Research Program to R.W., and by USPHS grant GM-1752 to E.B. J.D.C. is a recipient of an NIH Research Career Development Award 6-K4-GM-35, 190-01.

1. Van Dyke, H. B., Chow, B. F., Greep, R. O. & Rothen, A. (1942) *J. Pharmacol. Exptl. Ther.* **74**, 190–209.
2. Chauvet, J., Lenci, M. T. & Acher, R. (1960) *Biochim. Biophys. Acta.* **38**, 266–272.
3. Sawyer, W. H. (1961) *Pharmacol. Rev.* **13**, 225–277.
4. Walter, R., Schlesinger, D., Schwartz, I. L. & Capra, J. D. (1971) *Biochem. Biophys. Res. Commun.* **44**, 293–298.
5. Schlesinger, D. H., Walter, R., Schwartz, I. & Capra, J. D. (1971) *Fed. Proc.* **30**, 363.
6. Wuu, T. C., Crumm, S. & Saffran, M. (1971) *J. Biol. Chem.* **246**, 6043–6063.
7. Hollenberg, M. D. & Hope, D. B. (1968) *Biochem. J.* **106**, 557–564.
8. Breslow, E., Aanning, H. L., Abrash, L. & Schmir, M. (1971) *J. Biol. Chem.* **246**, 5179–5188.
9. Burford, G. D., Ginsburg, M. & Thomas, P. J. (1971) *Biochim. Biophys. Acta.* **229**, 730–738.
10. Breslow, E. & Walter, R. *Mol. Pharmacol.*, in press.
11. Dean, C. R., Hope, D. B. & Kazic, T. (1968) *Brit. J. Pharmacol. Chemother.* **34**, 192P–193P.
12. Fleischman, J. B., Porter, R. R. & Press, E. M. (1963) *Biochem. J.* **88**, 220–228.
13. Edman, P. & Begg, G. (1967) *Eur. J. Biochem.* **1**, 80–91.
14. Capra, J. D. & Kunkel, H. G. (1970) *Proc. Nat. Acad. Sci. USA* **67**, 87–92.
15. Hermodson, M. A., Ericsson, L. H. & Walsh, K. A. (1970) *Fed. Proc.* **29**, 728.
16. Pisano, J. J. & Bronzert, T. J. (1969) *J. Biol. Chem.* **244**, 5597–5607.
17. Van Orden, H. O. & Carpenter, F. H. (1964) *Biochem. Biophys. Res. Commun.* **14**, 399–403.
18. Edman, P. & Sjoquist, J. (1956) *Acta Chem. Scand.* **10**, 1507–1509.
19. Evans, J. V., King, J. W. B., Cohen, B. L., Harris, H. & Warren, F. L. *Nature* **178**, 849–850.
20. Boyer, S. H., Hathaway, P., Pascasio, F., Bordley, J., Orton, C. & Naughton, M. A. (1967) *J. Biol. Chem.* **242**, 2211–2232.
21. Bangham, A. D. (1957) *Nature* **179**, 467–468.
22. Schroeder, W. A., Shelton, J. R., Shelton, J. B., Robberson, B. & Babin, D. R. (1967) *Arch. Biochem. Biophys.* **120**, 124–135.
23. Walsh, K. A., Ericsson, L. H. & Neurath, H. (1966) *Proc. Nat. Acad. Sci. USA* **56**, 1339–1344.
24. Petra, P. H., Bradshaw, R. A., Walsh, K. A. & Neurath, H. (1969) *Biochemistry* **8**, 2762–2768.
25. Bearn, A. G. & Kunkel, H. G. (1968) *Textbook of Immunopathology* (Grune and Stratton, New York), pp. 48–59.
26. Niall, H. D., Hogan, M. L., Sauer, R., Rosenblum, I. Y. & Greenwood, F. C. (1971) *Proc. Nat. Acad. Sci. USA* **68**, 866–869.
27. Acher, R. (1966) *Angew. Chem. Int. Ed.*, **5**, 798–806.
28. Vliegenthart, J. F. G. & Versteeg, D. H. G. (1967) *J. Endocrinol.* **38**, 3–12.
29. Pink, J. R. L., Buttery, S. H., DeVries, G. M. & Milstein, C. (1970) *Biochem. J.* **117**, 33–47.