AMINO ACID SEQUENCE AND HETEROGENEITY OF GASTRIC INHIBITORY POLYPEPTIDE (GIP)

Hans JÖRNVALL, Mats CARLQUIST, Sam KWAUK, Susan C. OTTE, Christopher H. S. McINTOSH, John C. BROWN and Viktor MUTT

Departments of Chemistry I and Biochemistry II, Karolinska Institutet, S-104 01 Stockholm 60, Sweden and *Department of Physiology, University of British Columbia, Vancouver, V6T 1W5, Canada*

Received 18 November 1980

1. Introduction

Gastric inhibitory polypeptide (GIP) is one of a group of polypeptides with distant structural relationships to the glucagon–secretin family [1–3]. It has several effects but two major physiological activities have been recognized [4]. One is inhibition of gastric acid secretion (enterogastrone effect), explaining the present name and the bioassay followed in the initial purification [5] of the peptide. The other is stimulation of insulin release when administered in association with hyperglycemia (insulinotropic effect).

A 43-residue amino acid sequence for the polypeptide was reported in 1971 [1]. Based on this structure, synthetic replicates have been prepared to study the biological activity further. The same scheme has been followed with other gastrointestinal hormones and has usually produced a biologically fully active synthetic analogue. A recent example of this verification of a new hormone structure [6] is the finding that the synthetic form [7] of the mammalian gastrin-releasing peptide is active. However, in the case of GIP, comparisons between natural and synthetic material have not unequivocally confirmed the structure [4]. Synthetic products appear to contain only part of the bioactivity and immunoreactivity [8]. In addition, GIP preparations may contain minor components [4]. For these reasons, it was decided to re-investigate the primary structure of natural GIP preparations.

The results of a new sequence determination of GIP are now reported. It is shown that the previously suggested structure apparently is too long by a glutamine residue after position 29. The structure of the main component in GIP preparations described lacks this residue and is 42 residues long. It is in agreement with the total composition of native GIP, with results of repeated sequence degradations and with the properties of smaller fragments that have also been analyzed.

In addition, the present determination confirms that the GIP preparation is heterogeneous, and shows that the major secondary component appears to be identical to the most abundant component except that it lacks the first two residues of the latter. The second peptide is therefore suggested to be composed of residues 3–42 of the major component in the GIP preparation. Enzymatic processing as an explanation for the origin of the second peptide is discussed.

2. Materials and methods

2.1. *GIP*

The active material was extracted from the first part of pig upper intestine and purified as in [1,4,5]. The biological activity was followed by measuring the acid inhibitory effects [4]. Analysis by high-performance liquid chromatography (HPLC) was done in a Waters instrument on a μ-Bondapak C_{18} column using 28% acetonitrile, 0.1% phosphoric acid, 71.9% water as eluant, and detection at 215 nm.

2.2. *Structural analysis*

GIP (~100 nmol) was cleaved with CNBr (0.1 g) in 70% formic acid (0.4 ml) at room temperature for 24 h. A similar sample was cleaved at the two tryptophan residues in 30% acetic acid/5 M urea with *N*-chlorosuccinimide (720 μg, added twice with incubation for 1 h at room temperature after each addition). Excess reagent was removed by *N*-acetyl-

methionine (3.5 mg). This scheme follows [9] except for a larger excess of *N*-chlorosuccinimide (50-fold over tryptophan). Peptides obtained were purified by HPLC in a Waters instrument with a μ-Bondapak C_{18} column as above, but with a linear gradient of ethanol in an ammonium acetate–acetic acid system. This was composed of 5 mM ammonium acetate, 0.1% acetic acid, 1% ethanol (solvent A), while the gradient mixing solvent B had 89.91% ethanol, 10% solvent A and 0.09% acetic acid. Cleavage of GIP (~40 nmol) with trypsin (1:10, w/w) was performed in 0.1 M ammonium bicarbonate (0.2 ml) at 37°C for 4 h.

Amino acids were analyzed with a Beckman 121M analyzer. Samples were hydrolyzed in evacuated tubes for 22 h at 110°C with 6 M HCl/0.5% phenol, or for tryptophan determination, with 4 M methane sulfonic acid/0.2% 3-(2-aminoethyl)indole (Pierce). The dansyl-Edman method was used for manual sequence analysis with identification of dansyl-amino acids by thin-layer chromatography (TLC) on polyamide layers in 4 solvent systems [10]. Liquid-phase sequencer degradations in a Beckman 890C sequencer were performed with a 0.1 M quadrol peptide program in the presence of pretreated polybrene, and with identification of phenylthiohydantoin derivatives by HPLC (a Hewlett-Packard 1084B instrument) and by TLC, as in [11].

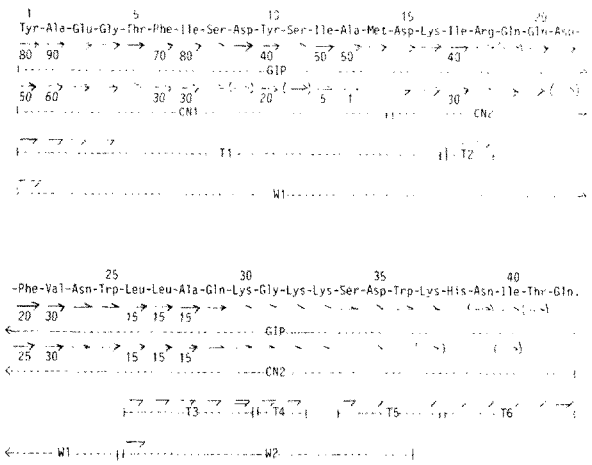

Fig.1. Amino acid sequence analysis of GIP, and the structure deduced for the major component. Analysis of GIP (~120 nmol) and its two CNBr fragments CN1 + CN2 (~100 nmol in mixture) by liquid–phase sequencer degradations are shown by –> for residues identified by HPLC (values give recoveries in nmol of stable thiohydantoin derivatives), and by ⟶ for residues also identified by TLC. Repetitive yields in the degradations: GIP 94% (Ala 2 –> 13), CN1 91% (Tyr 1 –> 10) and CN2 92% (Ile 17 –> Leu 27). Analysis of tryptic peptides (T1–T6 in mixture) by manual dansyl-Edman degradations are shown by —/, and recovery of the dansyl-amino acid also without hydrolysis, proving the C-terminal position is a peptide by ⟹. Peptide W1 and W2 are derived from *N*-chlorosuccinimide cleavages at tryptophan residues.

3. Results

3.1. *Structure of the main GIP component*

Direct liquid-phase sequencer analysis of the intact peptide revealed most of the structure up to residue 41, as given in fig.1. This analysis showed that the structure in [1] was largely correct, but also suggested that there should be 1 instead of 2 glutamine residues at positions 29–30. For confirmation, GIP was cleaved with CNBr. The peptide mixture obtained was submitted to liquid-phase sequencer degradation, without separation in order not to miss any fragments. Two major sequences, as expected from the presence of one methionine in the native peptide, were followed for 21 steps (CN1 and CN2, fig.1), in agreement with the first degradation.

Manual dansyl-Edman analyses were similarly performed on a tryptic digest of GIP. Several residues were detected in each cycle and could be ascribed to the expected fragments T1–T6 (fig.1). Residues 19–25 were not recovered, probably due to cyclization of the N-terminal glutamine, as noticed before in tryptic digests [11]. Instead, the region 26–30 was recovered (T3, fig.1), due to a chymotryptic-like cleavage, which is also common [12]. After a few cycles of degradation of the mixture, only peptides T3 and T6 (fig.1) remained in strong yield (dipeptides T2 and T4 had ended; T1 and T5 gave low sequence yields due to size and tryptophan, respectively). The dansyl results confirmed the sequencer analyses, the C-terminal region (peptide T6), and the C-terminal glutamine (fig.1).

The structure was further confirmed by cleavage of GIP with *N*-chlorosuccinimide at tryptophan residues and separation of the fragments by HPLC, as shown in fig.2. The total composition of fragment W2 (fig.1), corresponding to peak 1 in fig.2, is shown in table 1, and confirms the presence of only 1 glutamine residue at position 29–30. The entire sequence determination is further supported by the total composition of the whole native GIP (also shown in table 1). Consequently, based on these analyses, the primary structure of GIP is deduced to be as given in fig.1.

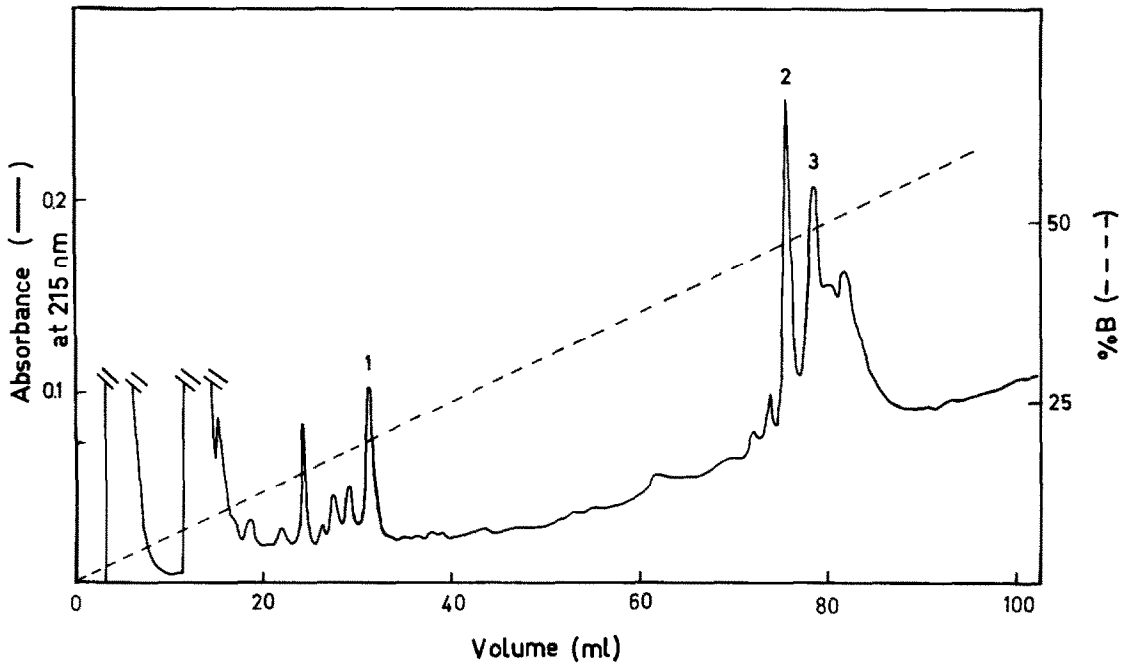

Fig.2. Separation of GIP fragments obtained by cleavage with *N*-chlorosuccinimide at tryptophan residues, HPLC on μ-Bondapak C_{18} with a gradient (as indicated) of solvent B (89.91% ethanol, 10% solvent A and 0.09% acetic acid) in solvent A (5mM ammonium acetate, 0.1% acetic acid, 1% ethanol). Flow rate 2 ml/min. The two large peaks eluted just after the start of the gradient are due to acetic acid and reagents from the cleavage. The numbered peaks 1,2 and 3 correspond to W2, W1 and W1A, respectively, in fig.1 and table 2.

3.2. *Structure of a second component in GIP preparations.*

The sequencer degradations of the intact fragment and of the CNBr peptide mixture revealed a contaminating sequence in low yield. This sequence was, in all positions that could be controlled (i.e., 35 of the first 38 residues), identical to the major GIP sequence except for starting at position 3 in the major structure. The uncorrected recoveries, shown in table 2, from the chromatograph values vary a little due to different backgrounds and chromatograph sensitivities, but average ~20–25% (table 2). The results therefore suggest that the second component in GIP preparations is present in a yield of ~20% relative to the major component, and that it is a fragment composed of residues 3–42. No evidence was obtained for any corresponding heterogeneity in the C-terminal region, but the N-terminal fragment W1 (fig.1) from the tryptophan cleavage was also recovered in two forms (corresponding to peaks 2 and 3 in fig.2). These two forms need not be equivalent to pure fragments of the two N-terminal components in GIP preparations, but could also contain secondary differences due to the pretreatments (e.g., oxidations or desamidations), but their total compositions were indistinguishable, except for the first 2 residues in GIP, tyrosine and alanine, which show lower values in the minor peak (table 1). The presence of a main GIP contaminant in ~20% yield, compatible with the sequencer results, is confirmed by direct fractionation of the GIP preparation on HPLC (fig.3).

4. Discussion

4.1. *Structure of GIP*

The amino acid sequence of GIP was determined by a combination of 4 different investigations, analysis of the intact peptide, analyses of 2 sets of fragments produced by chemical cleavages, and analysis of the tryptic peptides. Although the presence of a 2-residue shifted contaminant and of alternating lysine and glycine residues in the critical region (around position 30) complicated interpretations, the results suggest that the major component in GIP preparations is a 42-residue peptide. All data from sequence degrada-

Table 1
Total compositions of the GIP preparation investigated, and of the 2 large peptides after cleavage at tryptophan residues with *N*-chlorosuccinimide

Peptide composition	GIP	W1[a]	W2[a]
Asx	5.6 (6)	4.1 (4)	1.0 (1)
Thr	1.9 (2)	0.9 (1)	– –
Ser	3.2 (3)	1.9 (2)	0.9 (1)
Glx	5.2 (5)	3.0 (3)	1.1 (1)
Gly	2.3 (2)	1.2 (1)	1.1 (1)
Ala	3.1 (3)	1.9 (2)	1.0 (1)
Val	1.0 (1)	1.0 (1)	– –
Met	1.2 (1)	0.7 (1)	– –
Ile	3.5 (4)	2.7 (3)	– –
Leu	2.3 (2)	– –	1.8 (2)
Tyr	2.2 (2)	1.6 (2)	– –
Phe	2.1 (2)	1.7 (2)	– –
Trp	1.6 (2)	(1)	(1)
Lys	4.6 (5)	1.2 (1)	2.7 (3)
His	1.1 (1)	– –	– –
Arg	1.2 (1)	1.0 (1)	– –
Sum	42	25	11

[a] Composition of W1A identical within one decimal digit for all residues except Ala (1.7), Tyr (1.3) and Meo (0.4)

Values shown are molar ratios (values <0.2 omitted) without corrections for impurities or hydrolytic destructions, and with numbers from the sequence analyses within parentheses. Hydrolysis for 22 h with 6 M HCl except for tryptophan in GIP which was quantitated after hydrolysis with 4 M methane sulfonic acid. Tryptophan destroyed in W1 and W2, as well as methionine in W1, where it was partly recovered as the sulfone. Positions of peptides W1 and W2 are given in fig.1, peptide W1A is a variant of W1, and purification of all 3 peptides are shown in fig.2

tions and from total compositions of the intact preparation and its proteolytic fragments support the structure deduced. This sequence is one internal glutamine residue shorter than suggested in [1]. Significantly, it should be noted, however, that the corresponding region was previously recovered in low yield and was very difficult to analyze [1]. Furthermore, total compositions reported in [13] also agree with this analysis (table 1) as well as or better than with the previous, longer structure. All results therefore suggest that the major peptide in GIP preparations has the structure shown in fig.1.

4.2. *Heterogeneity of GIP preparations*

Results from both the fractionation (fig.3) and the sequence analysis (table 2) of the GIP preparation investigated show that the material is heterogeneous, containing a second component, corresponding to ~20% of the major component. A hetereogeneity

Table 2
Sequence analysis of second component in GIP preparations

Cycle	Residue identified	% of main component from degradation of	
		GIP prep.	CNBr peptide mix.
1	Glu	25	20
2	Gly	15	25
3	Thr	15	25
4	Phe	25	20
5	Ile	30	30
6	Ser	10	–
7	Asp	20	–
8	Tyr	25	15
9	Ser	20	–
10	Ile	35	25
11	Ala	10	20
12	Met	30	–
13	Asp	20	Average
14	Lys	25	recovery: 23%
15	Ile	20	
16	(Arg)	–	
17	Gln	25	
18	Gln	10	
19	Asp	20	
20	Phe	35	
21	Val	20	
22	Asn	15	
23	Trp	30	
24	Leu	10	
25	Leu	15	
26	Ala	35	
27	Gln	15	
28	Lys	25	
29	Gly	25	
30	Lys	30	
31	Lys	25	
32	(Ser)	–	
33	Asp	20	
34	Trp	20	
35	Lys	25	
36	(His)	–	
37	Asn	20	
38	Ile	10	
Average recovery		21%	

The amount of each residue is given as % recovered in relation to the recovery of the same residue 2 cycles later in the main component. Degradations in a liquid-phase sequencer and quantitation in a high-performance liquid chromatograph

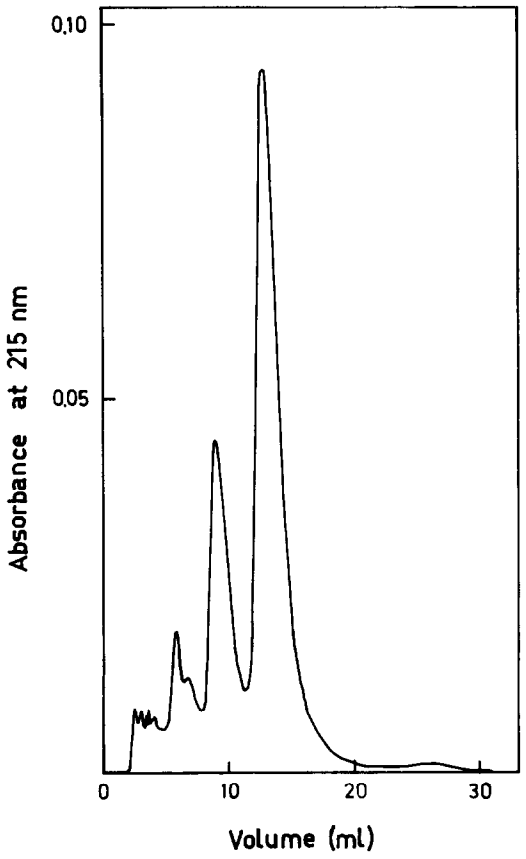

Fig.3. Separation of the GIP preparation by HPLC on μ-Bondapak C_{18} in 0.1% phosphoric acid, 28% acetonitrile, 71.9% water. Flow rate: 1.5 ml/min. A main component and a major second component are clearly seen apart from a minor third fraction.

had been reported, but the minor component was then reported to be present in lower yield, 5% [4] than the present second component. It appears possible that the degree of heterogeneity can vary between preparations, or that the third component in fig.3 was that previously detected. The structure of the second peptide is identical to residues 3–42 of the main form. The extensive similarity between the two components explains why they do not separate easily during the purification steps performed, and why a contamination of up to 20% or more, is still not visible in the total composition, as evidenced by the agreement between hydrolytic values (table 1; [13]) and the sequence of the main component (fig.1). An N-terminal heterogeneity with peptide chains that have not been separately purified, has also been found in preparations of bovine somatotropin [14–16].

The presence of a shorter but otherwise apparently identical second component raises questions on its origin and bioactivity. Regarding the origin, the structure of the deviating part, Tyr–Ala, is not at all reminiscent of the longer and C-terminally dibasic structures usually removed on regular processings of prohormone forms (cf, table 2 in [17]). Consequently, it does not appear likely that the two peptides in the GIP preparations represent usual prohormone–hormone forms. Instead, one possibility is that they are derived from different partly duplicated genes or different regions of a larger gene, if they are not just allelic variants. Deletion mutants have been detected in human somatotropin [18]. A genetic multiplicity for GIP would be compatible with the already large group of peptides related in that way [2,3] and with the multiple genetic origin of structures for hormones like melanotropin [19]. However, a different and perhaps more likely possibility is that the two forms in the GIP preparation are derived from secondary processings (or even degradations) through susceptibility to attack by aminopeptidase, elastase, dipeptidyl aminopeptidase or related enzymes in the intestine. It is even possible that the third component visible in fig.3 is one additional, processed form. Indeed, in view of the biological origin of GIP and other intestinal hormones, such an explanation for the origin of the second component in GIP preparations would appear less surprising than the facts that heterogeneities are not more frequently found among gastrointestinal peptides. A dipeptidyl peptidase has recently also been implied in another peptide conversion (of promelittin in the bee, [20]).

Regarding the bioactivity of GIP preparations, the discovery of two related peptides will necessitate further physiological studies. The main GIP component is already known to be active [4], but if this relates to both the main activities (enterogastrone effect and insulinotropic activity) and all other activities is unknown (cf. [4]). The possible activity of the second component is unknown but a biological activity cannot be excluded in view of the significant but reduced activity even of internally elongated synthetic analogues [8]. If the shorter form has altered bioactivities, wider structure–function relationships can be deduced from further studies of its properties.

Independent of future results on the origin and activity of both GIP components, the present results indicate one explanation of previous failures with synthetic analogues, show a new structure for GIP, and

establish the nature of a minor peptide constituent in the preparation.

Acknowledgements

This work was supported by grants from the Swedish Medical Research Council (projects 13X-3532, 13X-1010), the Swedish Cancer Society (project 620), Knut and Alice Wallenberg's Foundation and the British Columbia Health Care Research Foundation.

References

[1] Brown, J. C. and Dryburgh, J. R. (1971) Can. J. Biochem. 49, 867–872.
[2] Dayhoff, M. O. (1976) Atlas of Protein Sequence and Structure, vol. 5, suppl. 2, p. 125, National Biomedical Research Foundation, Silver Springs MD.
[3] Jörnvall, H., Mutt, V. and Persson, M. (1981) submitted.
[4] Brown, J. C., Dahl, M., Kwauk, S., McIntosh, C. H. S., Müller, M., Otte, S. C and Pederson, R. A. (1981) in: Gut Hormones, 2nd edn (Bloom, S. R. and Polak, J. M. eds) Churchill Livingstone, Edinburgh, in press.
[5] Brown, J. C., Mutt, V. and Pederson, R. A. (1970) J. Physiol. 209, 57–64.
[6] McDonald, T. J., Jörnvall, H., Nilsson, G., Vagne, M., Ghatei, M., Bloom, S. R. and Mutt, V. (1979) Biochem. Biophys. Res. Commun. 90, 227–233.
[7] Brown, M., Märki, W. and Rivier, J. (1980) Life Sci. 27, 125–128.
[8] Brown, J. C., Dryburgh, J. R., Frost, J. L., Otte, S C. and Pederson, R. A. (1978) in: Gut Hormones (Bloom, S. R. ed) pp. 277–282, Churchill Livingstone, Edinburgh.
[9] Lischwe, M. A. and Sung, M. T. (1977) J. Biol. Chem. 252, 4976–4980.
[10] Jornvall, H. (1970) Eur. J. Biochem. 14, 521–534.
[11] Jörnvall, H. and Philipson, L. (1980) Eur. J. Biochem 104, 237–247.
[12] Jornvall, H (1977) Eur. J. Biochem. 72, 425–442.
[13] Brown, J. C. (1971) Can. J. Biochem. 49, 255–261.
[14] Wallis, M. (1969) FEBS Lett. 3, 118–120.
[15] Santomé, J. A., Dellacha, J. M., Paladini, A. C., Wolfenstein, C. E. M., Peña, C., Poskus, E., Daurat, S. T., Biscoglio, M. J., De Sesé, Z. M. M. and De Sanguesa, A. V. F. (1971) FEBS Lett. 16, 198–200.
[16] Wallis, M. (1973) FEBS Lett. 35, 11–14.
[17] Pradayrol, L., Jörnvall, H., Mutt, V. and Ribet, A. (1980) FEBS Lett. 109, 55–58.
[18] Lewis, U. J., Bonewald, L. F. and Lewis, L. J. (1980) Biochem. Biophys. Res. Commun. 92, 511–516.
[19] Nakanishi, S., Inoue, A., Kita, T., Nakamura, M., Chang, A. C. Y., Cohen, S. N. and Numa, S. (1979) Nature 278, 423–427.
[20] Kreil, G., Haiml, L. and Suchanek, G. (1980) Eur. J. Biochem. 111, 49–58.