Hypothesis

# GASTRIN AND THE TRANSFORMING PROTEIN OF POLYOMA VIRUS HAVE EVOLVED FROM A COMMON ANCESTOR

G. S. BALDWIN

*Ludwig Institute for Cancer Research, Melbourne Tumour Biology Unit, Post Office Royal Melbourne Hospital, VIC 3050, Australia*

Received 30 November 1981

## 1. Introduction

Gastrin and the transforming protein (middle T antigen) of polyoma virus are superficially dissimilar molecules. However, this paper notes a hitherto unreported structural relationship between the 2 proteins which supports the hypothesis that they have evolved from a common ancestor.

Gastrin [1] is best known for its stimulation of acid secretion by the stomach epithelium, and of enzyme secretion by the stomach and pancreas [2]. More recently, the hormone has been shown to stimulate cellular proliferation both in vivo and in vitro [3–5]. Several overlapping forms which vary in length from 14–34 amino acids have been isolated [2]. It seems probable that all of these fragments are produced by proteolysis of a much larger prohormone, since the mRNA for pig gastrin is ~620 nucleotides long [6]. While the common C-terminal tetrapeptide amide Trp–Met–Asp–Phe-$NH_2$ elicits all the effects of the larger molecules, biological potency increases almost 2 orders of magnitude with increasing chain length, and is maximal for gastrin 9 (numbering from the C-terminal) [7].

Polyoma virus is able to produce a wide range of tumours in rodents [8]. The early region of the polyoma virus genome codes for 3 proteins: small T ($M_r$ 23 000), middle T ($M_r$ 50 000) and large T ($M_r$ 88 000) antigens [9]. The small T and large T proteins are not required for transformation [10], but mutational alteration of middle T protein causes changes in transformation frequency [11–16]. The dramatic reductions caused by mutations near the presumed membrane binding site [14] suggest that membrane attachment of middle T protein [17] is essential for transformation.

## 2. Structural homology between gastrin and middle T protein

Human gastrin is structurally homologous to a region of middle T protein of polyoma virus ~100 residues from its C-terminus. Homology is especially apparent around the sole tyrosine residue of gastrin, where 8 of 11 amino acids are identical (fig.1). With one gap inserted in the gastrin sequence the position of 14 amino acids in 34 (41%) is identical. Since polyoma virus infects rodents only, sequence homology with rodent rather than human gastrin may be even stronger.

Statistical comparison of the sequence shown in fig.1 by the method in [26] (table 1) provided unequivocal evidence of an evolutionary relationship between the C-terminal 17 amino acids of gastrin and middle T protein. The probability of relationship between the N-terminal 17 amino acids of gastrin and middle T protein was considerably lower. (The gastrin sequence was compared in 2 sections because it was necessary to insert a gap after residue 17 to maximize homology.) However, the similarity in the number and position of proline residues in the N-terminal gastrin segment and middle T protein suggests that the conformational restraints imposed by this unusual amino acid may produce a greater similarity in tertiary structure than is indicated by the statistical sequence analysis. A control comparison of gastrin with the oncoproteins of Rous and Moloney sarcoma viruses did not detect any significant homology in the region of the modified tyrosine (table 1).

Similarities have been reported in the arrangement of 2 clusters of cysteine residues (Cys–X–Cys–X–X–Cys) in the common region of the polyoma small and middle T proteins, and in the $\alpha$ and $\beta$ subunits of

*Published by Elsevier/North-Holland Biomedical Press*
00145793/82/0000–0000/$02.75 

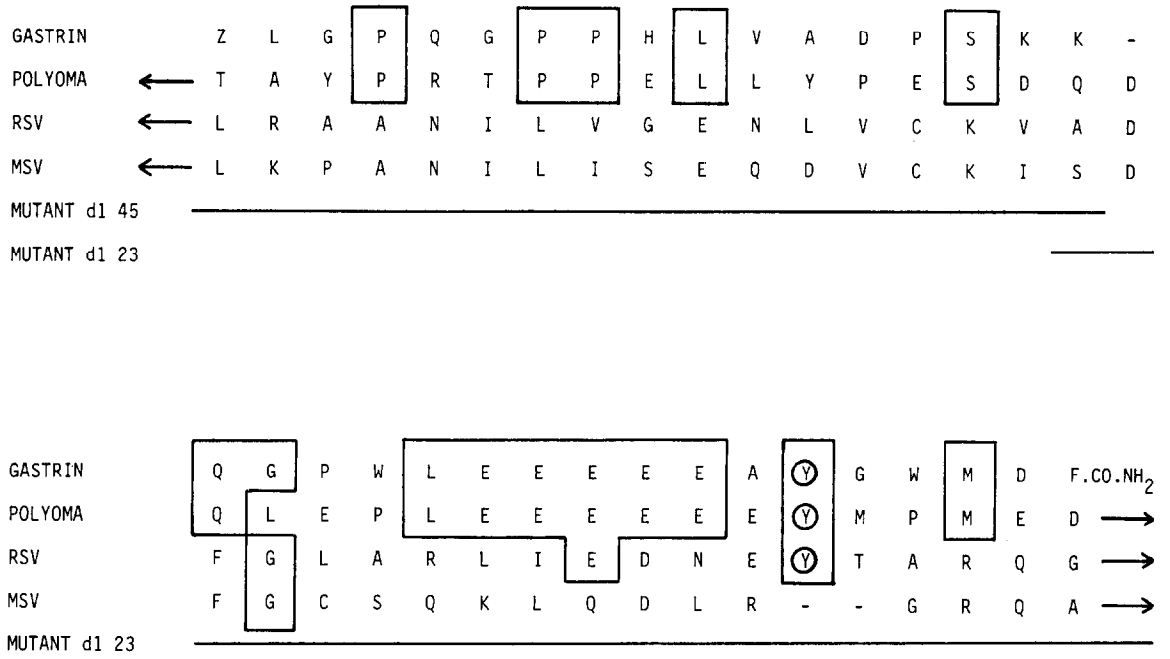


Fig.1. Comparison of the amino acid sequence of gastrin with polyoma middle T protein and other oncoproteins. The amino acid sequences of human gastrin [18,19], polyoma middle T protein [20,21], and the oncoproteins of Moloney [22] and Rous [23] sarcoma viruses (MSV and RSV, respectively) were aligned about their modified tyrosine residues, with one gap inserted in the gastrin sequence to maximize homology (see table 1). The C-terminus of gastrin is amidated [19]. Polyoma mutant dl 45 has 22 amino acids deleted, but transforms normally [12]. Polyoma mutant dl 23 has 34 amino acids deleted [24], transforms less efficiently than wild-type virus [11] (mean plating efficiency ± SEM: dl 23, 10.9 ± 11.1; wild-type, 65.7 ± 4.1 [13]) and has a considerably reduced protein kinase activity [25]. The one letter code for amino acids is: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine; Ⓨ modified tyrosine; Z, pyroglutamic acid. Amino acids identical to those in gastrin are enclosed in boxes. Arrows represent continuation of sequence, and deletions are underlined.

Table 1
Statistical comparison of the amino acid sequences of gastrin and viral oncoproteins [probability that alignment is a random occurrence (%)]

| Gastrin | Polyoma | RSV | MSV |
|---|---|---|---|
| N-terminal 17 residues | 10–50 | 50–95 | 95 |
| C-terminal 17 residues | 0.01– 0.01 | 50–95 | 50–95 |
| Entire molecule | 0.1 | 50–95 | 95 |

The alignments shown in fig.1 were compared by the method in [26]. Comparison of the uninterrupted gastrin sequence with the polyoma oncoprotein gave probability values of 50–95% for the N-terminal 17 residues, and 1% for the entire molecule

follitropin, lutropin, thyrotropin and chorionic gonadotropin [27,28]. However it seems improbable that there is any functional similarity between middle T protein and the hormones, since the same arrangement of cysteines is also found in two immunoglobulin γ chains, in one class of keratin, and in metallothionein [27,29]. Presumably the complete loss of transforming ability associated with deletion of both clusters in the polyoma host–range transformation mutant NG 18 [30] is only a reflection of the importance of disulphide bond formation by the cysteines in the maintenance of the active conformation of the protein.

The tyrosine residue is sulphated in ~1/2 of gastrin molecules [31]. The corresponding tyrosine residue of the middle T protein is phosphorylated by an associated protein kinase activity [32–34]. Since sulpho-

and phosphotyrosine are very similar in structure, the sequence homology between gastrin and polyoma virus middle T protein extends to homologous modification of a conserved tyrosine residue.

### 3. Origin of homology

The structural homology described in section 2 could have arisen by either convergent or divergent evolution. Thus 2 unrelated proteins may have independently converged to a similar sequence capable of recognition by related tyrosine sulpho- and phosphokinases. The preponderance of acidic amino acids on the N-terminal side of the modified tyrosines of gastrin, middle T protein, and the oncoprotein of Rous sarcoma virus may well represent such a recognition site. However, the fact that further homology is apparent only between gastrin and middle T protein strongly suggests that in this case the sequences have not arisen by convergent evolution. Similar arguments may be raised against the suggestion that the two proteins have independently evolved a polyglutamic acid sequence as a metal ion-binding site [35].

Alternatively the 2 proteins may have diverged from a common ancestor, in which case some vestigial functional similarities would be expected. Indeed the 2 proteins share the ability to promote cellular proliferation. Excessive production of gastrin in vivo by Zollinger-Ellison tumours is associated with hyperplasia of the fundic mucosa [36]. Conversely, reduction in gastrin levels following antrectomy results in parietal cell atrophy [37,38]. Both gastrin 17 and gastrin 34 have been shown to stimulate DNA synthesis in rat intestinal mucosa [39]. There may also be a connection between the increased incidence of gastric carcinoma [40] and the elevated serum gastrin levels [41] observed in patients with pernicious anaemia. Increased stem cell proliferation in the atrophic gastric mucosa in response to gastrin would be expected to increase DNA synthesis and thus the frequency of spontaneous neoplastic mutation. Viral transformation is of course accompanied by unrestrained cellular proliferation. The transforming function of middle T protein is inferred primarily from the transforming ability of the plasmid pPyMTI, which encodes only the middle T protein [10], and from deletion mutants such as dl 23 (fig.1), in which loss of a DNA sequence in the region coding for the conserved tyrosine residue results in a reduction in the efficiency of transformation [11,24]. Tumour formation in vivo occurs in a number of tissues including gastric epithelium, with the parotid gland as the favoured site [8]. This distribution may simply reflect differences in viral attachment to cell membranes rather than differences in the transformation frequency of infected cells in the various tissues. Thus the ability of both proteins to promote cellular proliferation appears to favour divergent evolution as the origin of the observed sequence homology.

A similar evolutionary relationship may well exist between epidermal growth factor (EGF) and the transforming growth factors (TGF) isolated from cells transformed by Moloney sarcoma virus [42] or by chemicals [43]. The observation that TGF binds to EGF receptors but not to EGF antibodies suggests limited structural homology [42]. However, comparison of the nucleotide sequence of the Moloney sarcoma virus oncogene [22] with the amino acid sequence of EGF [44] does not reveal any obvious similarities.

### 4. Mechanisms of transformation by polyoma virus

The fact that the DNA coding for the region of the polyoma middle T protein surrounding phosphotyrosine is not present in the related SV40 genome [20,21], suggests that polyoma virus may have evolved from an SV40-like virus by incorporation of rodent DNA sequences into its genome. Moreover, the sequence homology between gastrin and a region of the middle T protein suggests that the sequences included the rodent gastrin gene. Acquisition of the gastrin sequence would be expected to confer a significant selective advantage on the virus, since the trophic effects of the hormone would ensure a continual increase in the number of cells available for virus multiplication. Since middle T protein is membrane-bound [17] (and presumably intracellular in the intact cell), and since cell lysis is no longer required for viral transfer to unsaturated cells, both protein and virus would thus escape the animal's immune system. This mechanism is similar to that proposed in [45] for viral leukemogenesis, in which interaction of viral envelope antigens with surface receptors of infected lymphocytes 'stimulates the T cell to undergo continued rounds of antigen-induced proliferation'. However, several observations suggest that such a simple scheme, involving only the region of middle T pro-

tein homologous to gastrin, cannot be the only mechanism for papova virus transformation:

(i) The SV40 genome, which lacks the region homologous to the gastrin gene [20,21], can still transform cells efficiently;

(ii) The polyoma deletion mutant dl 23 which lacks most of the region of middle T protein homologous to gastrin can still transform cells, though with reduced efficiency compared to wild-type virus ([11,24], fig.1);

(iii) The plasmid pPyMTI, which encodes the middle T protein, but not the small T and large T proteins, is only 20–45% as efficient in transformation as the control plasmid containing the entire polyoma genome [10].

Thus polyoma virus appears to have evolved alternative transformation mechanisms [16].

The role of tyrosine modification in the transformation process is not clear. In the case of middle T protein some mutations which alter transformation frequency also alter the level of the associated protein kinase activity [25]. Although sulphation of tyrosine does not appear to alter the effectiveness of gastrin in stimulating either acid secretion [46] or DNA synthesis in rat intestinal mucosa [39], these results do not exclude the possibility that sulphate could be attached to, or removed from, the gastrin molecule within the target cell. However, the synthetic analogue pentagastrin [47], which has the structure T-butoxycarbonyl–β-Ala–Trp–Met–Asp–Phe–CO–$NH_2$, can elicit all the trophic effects of the parent molecule [3–5] (albeit at a significantly higher concentration [39]) even though it lacks the tyrosine residue. Further evidence that phosphotyrosine may not play an essential role in all transformation mechanisms comes from a comparison of the nucleotide sequences of the oncogenes of Moloney [22] and Rous [23] sarcoma viruses. Despite extensive homology the former sequence lacks the codon specifying the tyrosine residue whose autophosphorylation correlates with transformation [32,48].

## 5. Future experiments

The hypothesis that gastrin and the middle T protein of polyoma virus have evolved from a common ancestor suggests a number of experiments. In particular the mechanistic role of the modified tyrosine residue can be tested directly by insertion of polynucleotides coding for the C-terminal tetrapeptide of gastrin, and various extensions of it, into the appropriate region of the deletion mutant dl 23 genome to study their effects on transformation. The hypothesis is also consistent with the more general idea [49] that the viral oncoproteins responsible for tumorigenesis [10,50] have structurally homologous counterparts in uninfected cells [51,52], and that these 'cellular oncoproteins' are the growth regulators which control the proliferation of normal cells. Comparison of viral oncoprotein sequences with those of growth regulators like epidermal growth factor [44], nerve growth factor [53] and granulocyte–macrophage colony-stimulating factor [54] may therefore uncover further unexpected similarities.

## Acknowledgements

I thank colleagues too numerous to mention individually for their comments on this manuscript and in particular Dr Tony Burgess for first interesting me in the gastro-intestinal hormones.

## References

[1] Gregory, R. A. and Tracy, H. J. (1961) J. Physiol. 156, 523–543.
[2] Nilsson, G. (1980) in: Gastrointestinal Hormones (Jerzy-Glass, G. B. ed) Raven, New York.
[3] Johnson, L. R. (1976) Gastroenterology 70, 278–288.
[4] Johnson, L. R. (1977) Gastroenterology 72, 788–792.
[5] Johnson, L. R. (1980) in: Gastrointestinal Hormones (Jerzy-Glass, G. B. ed) Raven, New York.
[6] Noyes, B. E., Mevarech, M., Stein, R. and Agarwal, K. L. (1979) Proc. Natl. Acad. Sci. USA 76, 1770–1774.
[7] Tracy, H. J. and Gregory, R. A. (1964) Nature 204, 935–938.
[8] Stewart, S. E. (1960) Adv. Virus Res. 7, 61–90.
[9] Soeda, E., Arrand, J. R., Smolar, N., Walsh, J. E. and Griffin, B. E. (1980) Nature 283, 445–453.
[10] Treisman, R., Novak, U., Favaloro, J. and Kamen, R. (1981) Nature 292, 595–600.
[11] Griffin, B. E. and Maddock, C. (1979) J. Virol. 31, 645–656.
[12] Bendig, M. M., Thomas, T. and Folk, W. R. (1980) J. Virol. 33, 1215–1220.
[13] Ito, Y., Spurr, N. and Griffin, B. E. (1980) J. Virol. 35, 219–232.
[14] Novak, U. and Griffin, B. E. (1981) Nucleic Acids Res. 9, 2055–2071.
[15] Türler, H. (1980) Mol. Cell. Biochem. 32, 63–93.
[16] Griffin, B. E., Dilworth, S. M., Ito, Y. and Novak, U. (1980) Proc. R. Soc. London B 210, 465–476.

[17] Ito, Y., Brockelhurst, J. R. and Dulbecco, R. (1977) Proc. Natl. Acad. Sci. USA 74, 4666–4670.
[18] Gregory, R. A. and Tracy, H. J. (1975) in: Gastrointestinal Hormones (Thompson, J. C. ed) University of Texas Press, Austin, London.
[19] Choudhury, A. M., Kenner, G. W., Moore, S., Ramachandran, K. L., Thorpe, W. D., Ramage, R., Dockray, G. J., Gregory, R. A., Hood, L. and Hunkapiller, M. (1980) Hoppe-Seyler's Z. Physiol. Chem. 361, 1719–1733.
[20] Soeda, E., Arrand, J. R., Smolar, N. and Griffin, B. E. (1979) Cell 17, 357–370.
[21] Friedmann, T., Esty, A., LaPorte, P. and Deininger, P. (1979) Cell 17, 715–724.
[22] Van Beveren, C., Galleshaw, J. A., Jonas, V., Berns, A. J. M., Doolittle, R. F., Donoghue, D. J. and Verma, I. M. (1981) Nature 289, 258–262.
[23] Czernilofsky, A. P., Levinson, A. D., Varmus, H. E., Bishop, J. M., Tischer, E. and Goodman, H. M. (1980) Nature 287, 198–203.
[24] Smolar, N. and Griffin, B. E. (1981) J. Virol. 38, 958–967.
[25] Smith, A. E., Fried, M., Ito, Y., Spurr, N. and Smith, R. (1979) Cold Spring Harbor Symp. Quant. Biol. 44, 141–147.
[26] Moore, G. W. and Goodman, M. (1977) J. Molec. Evol. 9, 121–130.
[27] Friedmann, T., Doolittle, R. F. and Walter, G. (1978) Nature 274, 291–292.
[28] Ghosh, P. K., Piatak, M., Reddy, V. B., Swinscoe, J., Lebowitz, P. and Weissman, S. M. (1979) Cold Spring Harbor Symp. Quant. Biol. 44, 31–39.
[29] Dayhoff, M. O. (1972) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation Press, Washington DC.
[30] Soeda, E. and Griffin, B. E. (1978) Nature 276, 294–298.
[31] Rehfeld, J. F. (1980) in: Gastrointestinal Hormones (Jerzy-Glass, G. B. ed) Raven, New York.
[32] Smart, J. E., Opperman, H., Czernilofsky, A. P., Purchio, A. F., Erikson, R. L. and Bishop, J. M. (1981) Proc. Natl. Acad. Sci. USA 78, 6013–6017.
[33] Smith, A. E., Smith, R., Griffin, B. and Fried, M. (1979) Cell 18, 915–924.
[34] Eckhart, W., Hutchinson, M. A. and Hunter, T. (1979) Cell 18, 925–933.
[35] Palumbo, M., Jaeger, E., Knof, S., Peggion, E. and Wünsch, E. (1980) FEBS Lett. 119, 158–160.
[36] Neuberger, P. H., Lewin, M. and Benfils, S. (1972) Gastroenterology 63, 937–942.
[37] Lees, F. and Grandjean, L. C. (1958) Arch. Int. Med. 101, 943–951.
[38] Gjurldsen, S. T., Myren, J. and Fretheim, B. (1968) Scand. J. Gastroentcrol. 3, 465–470.
[39] Johnson, L. R. and Guthrie, P. D. (1976) Gastroenterology 71, 599–602.
[40] Hitchcock, C. R., MacLean, L. D. and Sullivan, W. A. (1957) J. Natl. Can. Inst. 18, 795–811.
[41] McGuigan, J. E. and Trudeau, W. L. (1970) New Eng. J. Med. 282, 358–361.
[42] DeLarco, J. E. and Todaro, G. J. (1978) Proc. Natl. Acad. Sci. USA 75, 4001–4005.
[43] Roberts, A. B., Lamb, L. C., Newton, D. L., Sporn, M. B., Delarco, J. E. and Todaro, G. J. (1980) Proc. Natl. Acad. Sci. USA 77, 3494–3498.
[44] Carpenter, G. and Cohen, S. (1979) Annu. Rev. Biochem. 48, 193–216.
[45] McGrath, M. S., Pillemer, E., Kooistra, D. and Weissmann, I. L. (1980) Contemp. Topics Immunobiol. 11, 157–184.
[46] Walsh, J. H., Debas, H. T. and Grossman, M. I. (1974) J. Clin. Invest. 54, 477–485.
[47] Morley, J. S., Tracy, H. J. and Gregory, R. A. (1965) Nature 207, 1356–1359.
[48] Levinson, A. D., Oppermann, H., Levintow, L., Varmus, H. E. and Bishop, J. M. (1978) Cell 15, 561–572.
[49] Bishop, J. M., Courtneidge, S. A., Levinson, A. D., Oppermann, H., Quintrell, N., Sheiness, D. K., Weiss, S. R. and Varmus, H. E. (1979) Cold Spring Harbor Symp. Quant. Biol. 44, 919–930.
[50] Sharp, P. A. (1979) Cold Spring Harbor Symp. Quant. Biol. 44, 1305–1322.
[51] Wyke, J. (1981) Nature 290, 629–630.
[52] Newmark, P. (1981) Nature 292, 15–16.
[53] Bradshaw, R. A. (1978) Annu. Rev. Biochem. 47, 191–216.
[54] Burgess, A. W. and Metcalf, D. (1980) Blood 56, 947–958.