

GENE 1631

## A remarkable amino acid sequence homology between a phage T4 tail fibre protein and ORF314 of phage $\lambda$ located in the tail operon

(Recombinant DNA; Gly-X-His-Y-His repeats; *E. coli* host; bacteriophage receptor; *Xenopus*)

Christian J. Michel<sup>a</sup>, Bernard Jacq<sup>\*\*</sup>, Didier G. Arquès<sup>b</sup> and Thomas A. Bickle<sup>\*\*\*</sup>

<sup>a</sup> Biozentrum, Basel University, Klingelbergstrasse 70, CH-4056 Basel (Switzerland) Tel. (061)25 38 80, and <sup>b</sup> Institut des Sciences Exactes et Appliquées, Université de Haute Alsace, 4 rue des Frères Lumière, F-68093 Mulhouse Cedex (France) Tel. 89 42 52 22

(Received March 3rd, 1986)

(Accepted April 17th, 1986)

---

### SUMMARY

We have found that the amino acid (aa) sequence of the tip of phage T4 tail fibre (gene 37) shows more than 50% homology with the aa sequence predicted from an open reading frame (ORF314) in the phage  $\lambda$  genome. ORF314 is near the 3' end of the late morphogenetic operon, beyond gene *J* coding for the  $\lambda$  tail fibre. The homologous sequences are for the most part composed of repeated aa, the most remarkable of which is a Gly-X-His-Y-His motif where X and Y are small, uncharged aa, found six times in the T4 protein and seven times in the  $\lambda$  ORF314 sequence.

---

### INTRODUCTION

The amount of protein and nucleic acid sequence data now available is so great that the computer has become an essential tool for its manipulation. A

comparison of these different sequences is expected to yield much new information on the rules governing biological structure and function. One simple example is the bias in codon usage that all organisms show (see, for example, Bennetzen and Hall, 1982).

We have been developing computer programs (to be described in detail elsewhere) that use combinatorial analysis to search for homologies in protein sequences. The programs allow repeated aa sequences to be found (in the same or different proteins) according to several user-defined parameters such as the percentage of identical aa in an analysis window of a specified length, and the minimum

---

\* Permanent address: Laboratoire de Génétique et Biologie Cellulaire, CNRS, Case 907, F-13288 Marseilles Cedex 9, (France) Tel. 91 82 05 78.

\*\* To whom correspondence and reprint requests should be addressed.

Abbreviations: aa, amino acid(s); gp, gene product; ORF, open reading frame.

|     |     |     |     |     |     |     |      |     |     |     |     |     |     |     |      |
|-----|-----|-----|-----|-----|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|------|
| Ser | Ser | Tyr | Pro | Ile | Gly | Ala | Pro  | Ile | Pro | Trp | Pro | Ser | Asp | Ser | 825  |
| Ser | Ala | Phe | Pro | Ala | Gly | Ala | Pro  | Ile | Pro | Trp | Pro | Ser | Asp | Ile | 82   |
| Val | Pro | Ala | Gly | Phe | Ala | Leu | Met  | Glu | Gly | Gln | Thr | Phe | Asp | Lys | 840  |
| Val | Pro | Ser | Gly | Tyr | Val | Leu | Met  | Gln | Gly | Gln | Ala | Phe | Asp | Lys | 97   |
| Ser | Ala | Tyr | Pro | Lys | Leu | Ala | Val  | Ala | Tyr | Pro | Ser | Gly | Val | Ile | 855  |
| Ser | Ala | Tyr | Pro | Lys | Leu | Ala | Val  | Ala | Tyr | Pro | Ser | Gly | Val | Leu | 112  |
| Pro | Asp | Met | Arg | Gly | Gln | Thr | Ile  | Lys | Gly | Lys | Pro | --- | Ser | Gly | 869  |
| Pro | Asp | Met | Arg | Gly | Trp | Thr | Ile  | Lys | Gly | Lys | Pro | Ala | Ser | Gly | 127  |
| Arg | Ala | Val | Leu | Ser | Ala | Glu | Ala  | Asp | Gly | Val | Lys | Ala | His | Ser | 884  |
| Arg | Ala | Val | Leu | Ser | Gln | Glu | Gln  | Asp | Gly | Ile | Lys | Ser | His | Thr | 142  |
| His | Ser | Ala | Ser | Ala | Ser | Ser | Thr  | Asp | Leu | Gly | Thr | Lys | Thr | Thr | 899  |
| His | Ser | Ala | Ser | Ala | Ser | Gly | Thr  | Asp | Leu | Gly | Thr | Lys | Thr | Thr | 157  |
| Ser | Ser | Phe | Asp | Tyr | Gly | Thr | Lys  | --- | --- | --- | --- | --- | --- | --- | 907  |
| Ser | Ser | Phe | Asp | Tyr | Gly | Thr | Lys  | Thr | Thr | Gly | Ser | Phe | Asp | Tyr | 172  |
| --- | --- | --- | Gly | Thr | Asn | Ser | Thr  | Gly | Gly | His | Thr | His | Ser | Gly | 919  |
| Gly | Thr | Lys | Ser | Thr | Asn | Asn | Thr  | Gly | Ala | His | Ala | His | Ser | Leu | 187  |
| Ser | Gly | Ser | Thr | Ser | Thr | Asn | Gly  | Glu | His | Ser | His | Tyr | Ile | Glu | 934  |
| Ser | Gly | Ser | Thr | Gly | Ala | Ala | Gly  | Ala | His | Ala | His | Thr | Ser | Gly | 202  |
| Ala | Trp | Asn | Gly | Thr | Gly | Val | Gly  | Gly | Asn | Lys | Met | Ser | Ser | Tyr | 949  |
| Leu | Arg | Met | Asn | Ser | Ser | Gly | Trp  | Ser | Gln | Tyr | Gly | Thr | Ala | Thr | 217  |
| Ala | Ile | Ser | Tyr | Arg | Ala | Gly | Gly  | Ser | Asn | --- | --- | --- | --- | --- | 959  |
| Ile | Thr | Gly | Ser | Leu | Ser | Thr | Val  | Lys | Gly | Thr | Ser | Thr | Gln | Gly | 232  |
| --- | --- | --- | --- | --- | --- | Thr | Asn  | Ala | Ala | Gly | Asn | His | Ser | His | 968  |
| Ile | Ala | Tyr | Leu | Ser | Lys | Thr | Asp  | Ser | Gln | Gly | Ser | His | Ser | His | 247  |
| Thr | Phe | Ser | Phe | Gly | Thr | Ser | Ser  | Ala | Gly | Asp | His | Ser | His | Ser | 983  |
| Ser | Leu | Ser | Gly | Thr | Ala | Val | Ser  | Ala | Gly | Ala | His | Ala | His | Thr | 262  |
| Val | Gly | Ile | Gly | Ala | His | Thr | His  | --- | --- | --- | --- | --- | --- | --- | 991  |
| Val | Gly | Ile | Gly | Ala | His | Gln | His  | Pro | Val | Val | Ile | Gly | Ala | His | 277  |
| --- | --- | Thr | Val | Ala | Ile | Gly | Ser  | His | Gly | His | Thr | Ile | Thr | Val | 1004 |
| Ala | His | Ser | Phe | Ser | Ile | Gly | Ser  | His | Gly | His | Thr | Ile | Thr | Val | 292  |
| Asn | Ser | Thr | Gly | Asn | Thr | Glu | Asn  | Thr | Val | Lys | Asn | Ile | Ala | Phe | 1019 |
| Asn | Ala | Ala | Gly | Asn | Ala | Glu | Asn  | Thr | Val | Lys | Asn | Ile | Ala | Phe | 307  |
| Asn | Tyr | Ile | Val | Arg | Leu | Ala | 1026 |     |     |     |     |     |     |     |      |
| Asn | Tyr | Ile | Val | Arg | Leu | Ala | 314  |     |     |     |     |     |     |     |      |

Fig. 1. A comparison of T4 gp37 and  $\lambda$  ORF314 proteins. The upper lines are the T4 sequence, the lower lines show those from  $\lambda$ . Identical aa residues are boxed. The numbering on the right margin corresponds to the position of the last aa in each line.

number of repeats. We have been working with bacteriophage protein sequences extracted from the National Biomedical Research Foundation (NBRF) protein sequence data bank. In the course of this work, we found that two of the phage protein sequences shared internal repeat sequences and a closer comparison of these sequences is the subject of this communication.

## EXPERIMENTAL

### (a) Amino acid homologies between T4 and $\lambda$ DNA

The homology that we have found is between the C-terminal fifth of the T4 tail fibre (gp37; Oliver and Crowther, 1981) and most of the aa predicted to be the product of ORF314 in the phage  $\lambda$  genome (Sanger et al., 1982). The ALIGN program of the Protein Identification Resource (NBRF, Washington, DC) gives a highly significant alignment score for the homology of 27 standard deviations using a gap penalty of one. This homology is shown in Fig. 1 with identical aa boxed. The homology begins at aa residue 811 in gp37 and at residue 68 in the ORF314 product and continues to the carboxyl terminus of both proteins. There are 141 identical aa, 65.6% of the gp37 sequence and 57.3% of the ORF314 product. The reason for the different percentages is that it was necessary to allow three gaps of 9–11 aa and a single aa deletion in the T4 sequence in order to make the alignment shown in Fig. 1. Two of the gaps can be accounted for by the peculiar repeated aa sequences found within these proteins.

### (b) Repeat sequences and the Gly-X-His-Y-His motif

The repeat sequences are shown in Fig. 2 (for the ORF314 product). One class of repeated sequence is 10 aa long and is found three times in the  $\lambda$  protein and twice in gp37, in both cases as contiguous, direct repeats (Fig. 2A). In the  $\lambda$  protein, two of these repeats are almost perfect (nine identical residues out of ten) while the third is less conserved. The second class of repeated sequence is 9 aa long (extended to 10 aa for two members) and is found seven times in the ORF314 product and six times in gp37 (Fig. 2B).

```

146 SerAlaSerGlyThrAspLeuGlyThrLys
156 ThrThrSerSerPheAspTyrGlyThrLys      A
166 ThrThrGlySerPheAspTyrGlyThrLys

176 SerThrAsnAsnThrGlyAlaHisAlaHis
190 SerThrGlyAlaAlaGlyAlaHisAlaHis
239 ThrAspSerGlnGlySerHisSerHis
253 AlaValSerAlaGlyAlaHisAlaHis      B
262 ThrValGlyIleGlyAlaHisGlnHis
271 ProValValIleGlyAlaHisAlaHis
280 SerPheSerIleGlySerHisGlyHis

      ThrValSerIleGlyAlaHisAlaHis  Consensus

```

Fig. 2. Repeated aa sequences in the ORF314 product. (A) The 10-aa sequence discussed in section b. (B) The 9-aa repeat. The aa shown in the consensus sequence are found in at least three of the copies, the underlined aa residues are present in all of them. The numbers refer to the first aa in each copy.

Some of the repeats are contiguous while others are separated by a few aa. The most remarkable feature of this repeat is the motif Gly-X-His-Y-His where X and Y are small, uncharged residues. This motif is found in all of the repeats in both the T4 and the  $\lambda$  protein.

Repeat units containing pairs of His and pairs of Cys residues separated from each other by two or four residues have recently been found in TFIIIA, a *Xenopus* transcription factor (Miller et al., 1985), where they are thought to be involved in binding divalent cations. Both  $\lambda$  and T4 require divalent cations for infection and it is possible that the phage repeats, though different from those in TFIIIA, are cation binding sites.

### (c) Relationship between T4 gp37 protein and $\lambda$ ORF314 product

The C-terminus of gp37 composes the tip of the phage tail fibre and contains the binding site for the phage receptor, lipopolysaccharide, in the outer membrane of its host, *Escherichia coli* (Wilson et al., 1970; Goldberg, 1983). There is no experimental evidence that ORF314 has a function. Indeed, the gene can be deleted or substituted as in  $\lambda plac5$  (Ippen et al., 1971; Daniels et al., 1983) and lytic growth on laboratory strains of *E. coli* is still possible. Here, we argue that several lines of evidence support the idea that ORF314 may be a gene whose product is involved in the interaction of the phage with its host.

The main  $\lambda$  receptor is the *lamB* gene product in the bacterial outer membrane which is recognised by the phage gene *J* product which is located at the tip of the tail (Randall-Hazelbauer and Schwartz, 1973; Murialdo and Siminovitch, 1972). However, there is some evidence that  $\lambda$  may employ an alternative to the *lamB*-*J* adsorption pathway under certain conditions. For example, it has been shown that  $\lambda$  can transduce strains with a deletion of the *lamB* gene at low efficiency (Braun-Breton and Hofnung, 1978) and can efficiently infect cells with the *lamB* gene repressed provided they are first starved in a  $Mg^{2+}$ -containing buffer (Arber et al., 1983).

ORF314 maps to the right of gene *J* and is oriented such that it would be transcribed with it (Sanger et al., 1982). The codon usage in ORF314 is typical of that used in other  $\lambda$  genes while that of gene *37* is typical for T4. This causes the aa sequences to be much more conserved than the corresponding nucleic acid sequences: of the 141 identical aa, 85 are coded by different codons, in eight cases by codons that differ by either two or three nucleotides. This shows that the sequences in  $\lambda$  and T4 have undergone considerable evolution since they shared a common ancestor and argues against the notion that one phage acquired the sequence from the other recently (in evolutionary terms).

#### (d) Conclusions

What could the ORF314 product be? One possibility would be a subunit of the short tail fibres (whiskers) seen by electron microscopy which have been assigned neither a gene nor a function (Katsura, 1983). These fibres are at the tip of the tail, ideally placed to interact with the cell surface during infection. We are currently testing this possibility.

#### ACKNOWLEDGEMENTS

We would like to thank H. Christen, E. Kellenberger, G. Lark, H. Moser, C. Nager and J. Shepherd for their interest and advice. Supported by Grants from the Swiss National Science Foundation and a Fellowship from the Centre National de la Recherche Scientifique to C.J.M.

#### REFERENCES

- Arber, W., Enquist, L., Hohn, B., Murray, N.E. and Murray, K.: Experimental methods for use with lambda. In Hendrix, R.W., Roberts, J.W., Stahl, F.W. and Weisberg, R.A. (Eds.), *Lambda II*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1983, pp. 433-466.
- Bennetzen, J.L. and Hall, B.D.: Codon selection in yeast. *J. Biol. Chem.* 257 (1982) 3026-3031.
- Braun-Breton, C. and Hofnung, M.: Explanations accounting for transduction by bacteriophage  $\lambda$  in maltose negative bacteriophage  $\lambda$  resistant mutants of *Escherichia coli* K-12. *Mol. Gen. Genet.* 159 (1978) 143-149.
- Daniels, D.L., Schroeder, J.L., Szybalski, W., Sanger, F. and Blattner, F.R.: A molecular map of coliphage lambda. In Hendrix, R.W., Roberts, J.W., Stahl, F.W. and Weisberg, R.A. (Eds.), *Lambda II*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1983, pp. 469-517.
- Daniels, D.L., Schroeder, J.L., Szybalski, W., Sanger, F., Coulson, A.R., Hong, G.-F., Hill, D.F., Petersen, G.B. and Blattner, F.R.: Complete annotated lambda sequence. In Hendrix, R.W., Roberts, J.W., Stahl, F.W. and Weisberg, R.A. (Eds.), *Lambda II*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1983, pp. 519-676.
- Goldberg, E.: Recognition, attachment and injection. In Mathews, C.K., Kutter, E.M., Mosig, G. and Berget, P.B. (Eds.), *Bacteriophage T4*. American Society for Microbiology, Washington, DC, 1983, pp. 32-39.
- Ippen, K., Shapiro, J.A. and Beckwith, J.R.: Transposition of the *lac* region to the *gal* region of the *Escherichia coli* chromosome: isolation of  $\lambda$ lac transducing bacteriophages. *J. Bacteriol.* 108 (1971) 5-9.
- Katsura, I.: Tail assembly and injection. In Hendrix, R.W., Roberts, J.W., Stahl, F.W. and Weisberg, R.A. (Eds.), *Lambda II*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1983, pp. 331-346.
- Miller, J., McLachlan, A.D. and Klug, A.: Repetitive zinc binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* 4 (1985) 1609-1614.
- Murialdo, H. and Siminovitch, L.: The morphogenesis of bacteriophage lambda, IV. Identification of gene products and control of the expression of the morphogenetic information. *Virology* 48 (1972) 785-823.
- Oliver, D.B. and Crowther, R.A.: DNA sequence of the tail fibre genes 36 and 37 of bacteriophage T4. *J. Mol. Biol.* 153 (1981) 545-568.
- Randall-Hazelbauer, L. and Schwartz, M.: Isolation of the bacteriophage lambda receptor from *Escherichia coli*. *J. Bacteriol.* 116 (1973) 1436-1446.
- Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F. and Petersen, G.B.: Nucleotide sequence of bacteriophage  $\lambda$  DNA. *J. Mol. Biol.* 162 (1982) 729-773.
- Wilson, J.H., Luftig, R.B. and Wood, W.B.: Interaction of bacteriophage T4 tail fiber components with a lipopolysaccharide fraction from *Escherichia coli*. *J. Mol. Biol.* 51 (1970) 423-434.

Communicated by K.F. Chater.