Volume 208, number 1 FEBS 4156 November 1986

The phage Mu repressor c and IS30 transposase proteins are significantly related

Brian Dalrymple

CSIRO Division of Molecular Biology, PO Box 184, North Ryde, NSW 2113, Australia

Received 1 September 1986

The IS30 transposase exhibits significant amino acid sequence homology to the phage Mu repressor c in the amino- and carboxy-terminal regions of the proteins. The conserved sequences include the proposed Mu repressor DNA binding site, which is also related to the proposed Mu and D108 transposase DNA binding sites. The carboxy-terminal homologies are characterised by two almost complete, and one partial, somewhat diverged amino acid sequence repeats. Only weak homologies to this domain are present in the Mu transposase (Mu A). Nevertheless, a clear link between an insertion sequence and a bacteriophage has been established.

Phage Mu *Insertion sequence IS30* *Tn3 Transposase* (Escherichia coli)

1. INTRODUCTION

We have previously reported that the terminal inverted repeats of the *Escherichia coli* insertion sequence IS30 exhibit DNA sequence homology to the ends of the phage Mu [1]. Like IS30, phage Mu replicates by transposition and the homologous bases correspond to those implicated in the binding of the Mu transposase (Mu A) at the ends of the phage [2,3]. The repressor of phage Mu transposition (Mu c) binds to sites which are related to the Mu A binding sites [2] and comparison of the sequences of Mu A and Mu c has identified a region of homology in the amino-terminal end of both proteins which has the potential to form the helix-turn-helix motif of many DNA binding proteins [4]. The Mu-like phage, D108, also encodes a transposase which binds to the same sites as Mu A and Mu c [2]. This transposase (D108 A) has extensive sequence homology with Mu A [4] and, although the amino-terminal sequences are not closely related, the proposed helix-turn-helix motif is retained [5]. Phage Mu itself also encodes another protein involved in transposition, Mu B (see [6]) which also has a similar amino-terminal helix-turn-helix motif [5].

In our previous analysis the amino acid sequence of the IS30 transposase was compared with the first 88 amino acids of Mu A, but not to Mu c [1]. In addition the approach taken was not sensitive enough to identify weak homologies. However, the data outlined above has prompted a more detailed reanalysis of the relationship of the IS30 transposase to phage Mu encoded proteins. The ends of transposon Tn3 also have a small region of homology to the sequence of the Mu A binding site [3] and the Tn3 transposase (tnp A) was also included in the analysis.

2. EXPERIMENTAL

The amino acid sequences of the long open reading frame (ORFA) of IS30 [1] and of Mu A [5], Mu c [7], Mu B [8], D108 A [4] and tnp A [9] were initially compared using dot matrix plots generated using the MDCOLOUR program (Reisner, A. and Bucholtz, C., unpublished) with a wide range of different parameters. Final alignment of the sequences was carried out by hand. The significance of the alignments was calculated using the SEQDP program [10]. This program aligns random sequences generated from one of

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

the proteins with the other and calculates the number of standard deviations away from the mean score of the random alignments that the actual alignment lies. The parameters of 100 random alignments, a deletion penalty of 4 and a bias of zero were used.

3. RESULTS AND DISCUSSION

3.1. *Amino-terminal homologies*

It has previously been shown that the strongest homologies between Mu c and the phage encoded transposases lie in the amino-terminal regions of the proteins [5]. The alignments presented in this paper confirm this observation (fig.1) and the statistical analyses show that Mu A and Mu c are the most similar of the three proteins in this region (table 1). The D108 A protein is somewhat less related to Mu A than Mu c is in the amino-terminus, but the rest of the two transposases are probably much more closely related than in this region [4]. However, the alignment of D108 A and Mu c over the first 70 or so amino acids is on the borderline of significance (table 1). The alignment of Mu c and the IS30 transposase in this region also exhibits a similar low level of significance

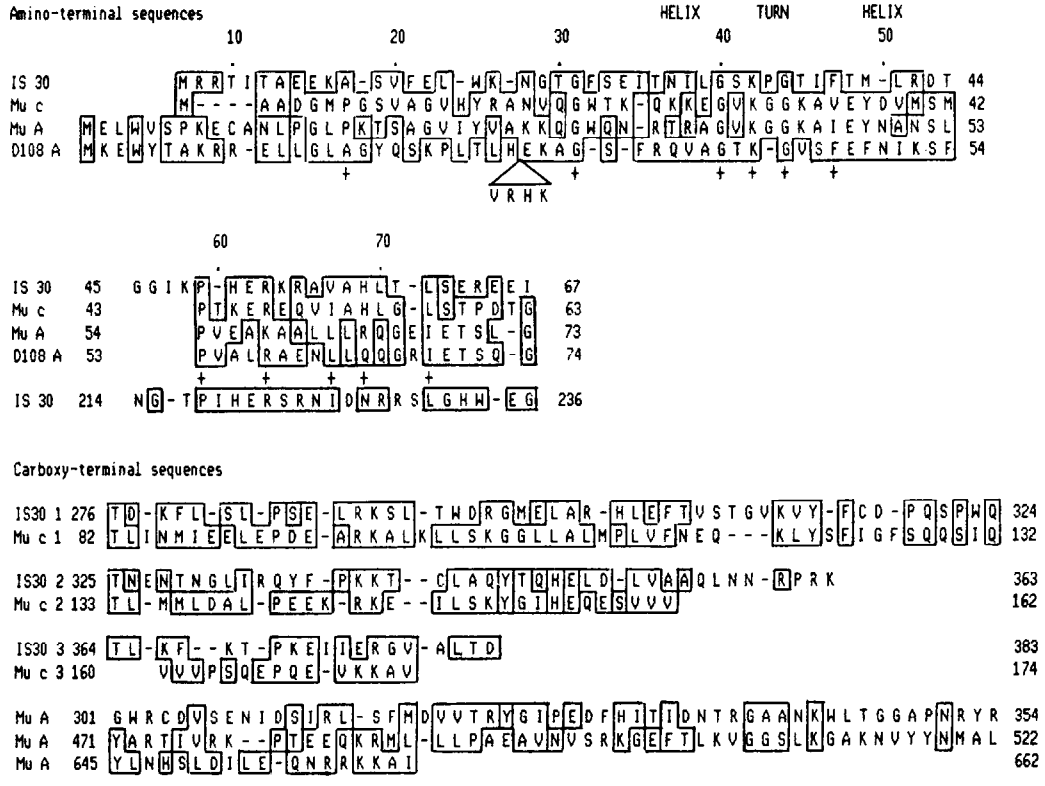

Fig.1. Alignments of the protein sequences. Identities and conservative substitutions are boxed and additional different sets of identities and conservative substitutions are in separate boxes. In the IS30 sequences between amino acids 214 and 236 the boxed amino acids are identical to the equivalent residue in one of the first four sequences or are conservative substitutions. Conservative substitutions are those substitutions which have positive scores in the 250 PAMs substitution matrix of Dayhoff et al. [11]. Deletions (−) were introduced to maximize the homologies in the alignment. Coordinates over the amino-terminal sequences are based on the total alignment, all other coordinates are for the relevant protein as indicated on the left hand side. The crosses under the amino-terminal sequences indicate identities or conservative substitutions present in all four proteins. The repeated sequences in the carboxy-terminal regions of Mu c and IS30 are numbered 1, 2 and 3. The proposed helix-turn-helix DNA binding site [5] is indicated above the sequences.

Table 1

The significance of the amino acid alignments

Protein	Amino acid no.	Mu c (1–63)	Mu A (1–73)	D108 A (1–72)	Mu B (1–70)	tnp A (1–70)	Mu c (82–174)	Mu A (74–184)
IS30	(1– 67)	2.2[a]	0.6	0.1	−0.2	−0.3		
Mu c	(1– 63)		10.6[a]	2.5[a]	−0.9	−0.8		
Mu A	(1– 73)			6[a]	1.3	−0.1		
D108 A	(1– 72)				1.8	−0.3		
Mu B	(1– 70)					0.7		
IS30	(276–383)						2.7[a]	0.3
Mu c	(64–174)							1.3

[a] Significant figures

Results expressed as number of standard deviations from the mean of random alignments

(fig.1); however in the light of the known relationship between D108 A and Mu A (and hence Mu c) it is probable that the observed alignment of the IS30 transposase and Mu c does reflect a true homology. This conclusion is also supported by the similar locations of the two regions in the proteins. In contrast, the alignments of the amino-terminal regions of Mu B and tnp A with the equivalent regions of the other proteins are clearly not significant (table 1).

The sequence homologies between Mu c and the IS30 transposase include the proposed DNA binding region (fig.1) within which glycine 40 (coordinates of the total amino acid alignment), lysine 42, glycine 44 and a hydrophobic amino acid at position 47 are conserved in all four proteins, suggesting that these are the most important amino acids in this region. Around amino acids 60–80 are a number of other amino acids which are also conserved between all of the proteins (fig.1). IS30 itself also has a second similar sequence around amino acid 225 (fig.1). The function of these domains is not known, although perhaps they are also involved in the DNA binding. The homology of the IS30 transposase to the proposed DNA binding site of the phage proteins strengthens the proposal [5] that the first region is involved in the binding. This also suggests that the IS30 transposase may well bind to these sequences in the terminal inverted repeats of IS30 which are homologous to the Mu A binding sites.

3.2. *Carboxy-terminal homologies*

In the preliminary dot matrix plots, in which the proteins were plotted against themselves, we observed that the carboxy-terminal region of the Mu c protein contained two almost complete and a third partial sequence repeat (fig.1). A number of amino acids at the junction of the second and third repeats could lie in either repeat. The IS30 transposase has sequences which are clearly homologous to the Mu c repeat sequences (fig.1) and overall the carboxy-terminal regions of the two proteins are probably significantly related (table 1). The overall alignment of the two proteins (omitting amino acids 64–81 from Mu c and 68–275 from the IS30 transposase) is more significant, being 4 standard deviations from the mean of random alignments. In contrast, the Mu A protein has no obvious repeated sequence nor strong homologies to this common repeated sequence, although a number of regions of Mu A, including the carboxy-terminal sequence itself, do have limited homology to the repeated region (fig.1). In addition, these analyses also suggest that the homology between the carboxy-terminal region of Mu c (amino acids 64–174) and the amino-terminal region of Mu A (amino acids 74–184) as proposed by Harshey et al. [5] is probably not significant (table 1).

Mu c and the IS30 transposase are apparently more closely related to each other than either is to Mu A. Indeed, most of the sequence of Mu c is homologous to sequences in the IS30 transposase, although the latter also contains additional amino acids. The repressor presumably does not have any transposase activity, thus it is probable that the sequences involved in the enzymatic activity of the

transposase lie in the unique central portion of the IS30 protein. However, these sequences do not appear to be closely related to those of Mu A which, although it is a much larger and presumably more complex protein than the IS30 transposase, is also a transposase. The function of the conserved carboxy-terminal domains is not known, but considering their presence in Mu c and the IS30 transposase and absence in Mu A they are probably not involved in DNA binding or in the enzymatic activity of the IS30 transposase. A possible function is in subunit-subunit interaction in multimers of these proteins, interactions which are perhaps different in, or not required by, Mu A.

The Mu A protein and the IS30 transposase may have evolved from a DNA binding protein similar to the present Mu c protein, either independently acquiring transposition activity, or rapidly evolving from a progenitor transposase. Alternatively, all three proteins could have evolved from an original transposase encoded by an insertion sequence-like element, or an ancestral Mu-like phage, followed by considerable change in protein structure and sequence as well as the nature of the element.

In conclusion the analyses indicate that the putative IS30 transposase is significantly related to the phage Mu repressor and that both proteins probably share a common ancestor. Thus a clear link between an insertion sequence and a bacteriophage is established.

ACKNOWLEDGEMENTS

I would like to thank John S. Mattick for helpful discussions and Anne McGill for typing the manuscript.

REFERENCES

[1] Dalrymple, B., Caspers, P. and Arber, W. (1984) EMBO J. 3, 2145–2149.
[2] Craigie, R., Mizuuchi, M. and Mizuuchi, K. (1984) Cell 39, 387–394.
[3] Groenen, M.A.M. and Van de Putte, P. (1986) J. Mol. Biol. 189, 597–602.
[4] Toussaint, A., Faelen, M., Desmet, L. and Allet, B. (1983) Mol. Gen. Genet. 190, 70–79.
[5] Harshey, R.M., Getzoff, E.D., Baldwin, D.L., Miller, J.L. and Chaconas, G. (1985) Proc. Natl. Acad. Sci. USA 82, 7676–7680.
[6] Toussaint, A. and Resibois, A. (1983) in: Mobile Genetic Elements (Shapiro, J.A. ed.) pp.105–158, Academic Press, New York.
[7] Priess, H., Kamp, D., Kahmann, R., Brauer, B. and Delius, H. (1982) Mol. Gen. Genet. 186, 315–321.
[8] Miller, J.L., Anderson, S.K., Fujita, D.J., Chaconas, G., Baldwin, D.L. and Harshey, R.M. (1984) Nucleic Acids Res. 12, 8627–8638.
[9] Heffron, F., McCarthy, B.J., Ohtsubo, H. and Ohtsubo, E. (1979) Cell 18, 1153–1163.
[10] Kanehisa, M., Klein, P., Greif, P. and DeLisi, C. (1984) Nucleic Acids Res. 12, 417–428.
[11] Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1978) in: Atlas of Protein Sequence and Structure (Dayhoff, M.O. ed.) vol.5, suppl.3, pp.345–352, Nat. Biomed. Res. Found., Washington, DC.