Volume 12 Number 10 1984 Nucleic Acids Research

# Rat prostatic steroid binding protein: characterisation of the Alu element upstream of the C3 genes

Helen C.Hurst and Malcolm G.Parker

Imperial Cancer Research Fund, P.O. Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK

Received 5 March 1984; Revised and Accepted 26 April 1984

ABSTRACT

We have characterised an Alu-like repeptive element found about 400 bp upstream of the gene encoding the C3 component of rat prostatic steroid binding protein and suggest, from comparisons with other published sequences, that it is an example of a third class of rodent Alu-equivalent sequences. Members of this class are 80-90 bp long, share greater than 90% sequence homology, and contain sequences resembling the RNA polymerase III bipartite promoter. The Alu type III element within the C3 gene was found to be expressed in cell-free systems and within heterologous cells stably transfected with the C3 gene, but these transcripts were not detectable in rat ventral prostate. It therefore seems unlikely that expression of this sequence plays a role in expression from the adjacent C3 gene.

## INTRODUCTION

The Alu family of middle repetitive DNA sequences is represented to the order of $10^5$ times within the genomes of primates and rodents (1). These sequences are interspersed among protein coding DNA sequences, as evidenced by the isolation of many examples of Alu repeats within the intervening and flanking DNA of cloned structural genes, a good example of this being within the human β globin gene cluster (2).

In the human genome, there may be only one type of Alu repeat (3), but in many species of rodents two different families have now been described (4), termed type 1 (similar to primate Alu) and type 2 Alu-equivalent sequences (1). Although distinct in length and sequence, these two families possess common structural features including: i) a precisely defined 5' end characterised by a consensus sequence for each family, whereby each individual member shows some 80-90% homology to the appropriate consensus; ii) an imprecisely defined, A-rich 3' end; iii) terminal flanking direct repeats of single copy cell DNA of 6-21 bp. The presence of the flanking repeats suggests that these elements may be mobile in the genome, but are not transposable elements per se (1). Alu repeats usually contain consensus sequences for the bipartite RNA poly-

merase III promoter (5) and have been shown in vitro (6) and in vivo (4) to be transcribed by this enzyme. Alu sequences may also be expressed as part of RNA polymerase II transcripts (7). Models have been proposed whereby the apparent mobility of Alu elements is explained by the return of their transcripts to the genome (8,9).

The accumulated information to date on Alu elements has therefore proved very intriguing, particularly in the light of their close physical relationship with protein coding genes. It has been suggested that DNA repeats, or their transcripts, may play a role in the control of expression from adjacent structural genes (10). Recent evidence comes from studies of one such class, referred to as identifier or ID sequences (11-13), which are detected primarily in brain RNA and may be involved in tissue specific gene expression. However, other theories state that such repeat sequences serve no useful function and their ability to show mobility within the genome allows them to escape natural selection and remain within the cell DNA as the ultimate parasite (14,15).

Obviously to assess the merits of these two viewpoints it is essential to investigate the importance of individual repeats found associated with particular genes. We have described the presence of a short Alu-like repetitive element (RE) referred to as Alu type III within the 5' flanking DNA of two non-allelic genes [C3(1) and C3(2)] encoding the C3 component of rat prostatic steroid binding protein. In this paper we show that this Alu sequence is probably a member of a third, distinct class of rodent Alu-equivalent repeats. Furthermore, in view of the close proximity of this Alu type III element to the start of C3 transcription, we analysed its expression, but failed to find a corollary between expression from C3 and the Alu sequence.

## MATERIALS AND METHODS

### Cell lines

The C3 transformed S115 cell lines were the gift of Dr. M. Page and were obtained using the dominant selection vector pSV2-gpt (17). L50 cells contained the entire C3(1) gene as a 9 Kb Bam H1 DNA fragment, at 50-100 gene copies/cell. The hormone-responsive cell lines SV2 and SV6 both possessed a 6 Kb Xba 1-Bam H1 fragment of C3(1) which included only $\sim$800 bp of 5' flanking DNA (17, and M. Page unpublished). Cultures of these cells were grown from frozen stocks in standard S115 medium (Eagle's 2% foetal calf serum, 30 mM unbuffered HEPES, 3 x $10^{-8}$M testosterone). Total cyto-

plasmic RNA and poly(A)-containing RNA from these cell lines and rat ventral prostate were prepared as previously described (17,18).

Primer extension and S1 mapping

Primers (ss) and probes (ds) were derived from subclones of the C3(1) gene. Protocols for primer extension were a previously described (16,20) and the products were run on 6% sequencing gels (20). Three day exposures of the gels were scanned using a Joyce Loeble chromoscan 3 densitometer.

In vitro transcription

A C3(1) subclone was treated with restriction enzymes to excise the Alu element. This DNA was assayed at 40 µg/ml within transcription mixes based on HeLa cell extracts (21) kindly provided by Dr. G. Veldman. The precise protocol for *in vitro* transcription and removal of DNA template was as described in Jat *et al*. (22). The RNA recovered was analysed by primer extension.

## RESULTS

THe Alu-like repetitive element upstream of the rat prostatic steroid binding protein C3 genes was first identified by sequence comparison with the E.M.B.O. sequence library. A short element located between -450 to -370 was identified (16) which shares 90-95% homology to several repetitive sequences already found in the rat genome or expressed in certain rat tissues (see Fig. 1). Clearly these sequences are 'Alu-like' as they possess the characteristic general structural features described in the Introduction and additionally have two restriction sites for the enzyme Alu I. Furthermore, all the genomic isolates were found to be flanked by direct repeats of cell DNA which were unique to the location of each particular repeat. The C3 example can also be assigned direct flanking DNA repeats, though their exact position is unclear as the genomic sequences flanking the repeat are themselves very repetitious. If repeats of 7 bp or more alone are considered, then three possible pairs have been indicated in Fig. 2.

Despite these structural similarities with known Alu sequences, computer analysis (not shown) revealed that these short ($\sim$ 80 bp) Alu-like elements only possess at most 62% homology to a Chinese hamster 166 bp type 1 sequence (a rat sequence has not been published) and 61% homology to a rat 190 bp type 2 element. Therefore it is suggested that the sequences represented in Fig. 1 are members of a distinct rodent type 3 Alu-equivalent family.


Fig. 1 A comparison of six short, Alu-like rat repetitive elements. The Alu-like sequence found between -450 and -370 in the C3 genes is given in full in line 1. Alu I restriction sites are boxed and the space in the sequence has been introduced to help align the other published examples represented underneath. A dot indicates these sequences are identical to line 1, while nucleotide differences are written in. Line 2 represents a sequence found within a rat α-tubulin pseudogene which itself otherwise resembled an insertion of mature α-tubulin mRNA into the genome (28). Line 3 represents a sequence found within the second intron of the rat growth hormone gene isolated by Barta et al. (24). The intron in fact contained three repetitive elements: a tandem repeat of a rodent type 2 Alu-equivalent sequence, followed by this shorter element. Lines 4 and 5 represent sequences found at the 3' ends of two cDNA clones isolated from rat brain p(A+) RNA by Sutcliffe et al. (4). The RE was the only common sequence between the two cDNA's and was also the only portion to hybridise to three other similarly isolated cDNA clones and to a 160 bp RNA found in brain, but not liver or kidney tissue. Line 6 represents a sequence producing an allelic variation approximately 8 Kb upstream of the rat prolactin gene (25).

The top line in the Figure shows the general dimensions of genes transcribed by RNA polymerase III (5). Consensus sequences for the split 'promoter', or polymerase binding regions, are given in full above matching sequences within the six Alu repeats.

As these type 3 repeats contain sequences that may be sufficient to render them an RNA polymerase template, it is conceivable that the C3 RE is expressed in vivo. This possibility is particularly intriguing in view of the close proximity of this element to the promoter regions of the C3 gene, whose expression is controlled by androgens (17,23). For this reason we examined the expression of these elements in rat ventral prostate, where the C3(1) gene is active, and also within cell lines produced by transfection of the C3 gene into androgen responsive mouse mammary carcinoma (S115) cells (17). By using gene transfer into mouse cells it was possible to examine expression of the Alu type III element associated with the C3 gene in the absence of other members of the family because mouse S115 cells do not produce detectable transcripts of this type (not shown).

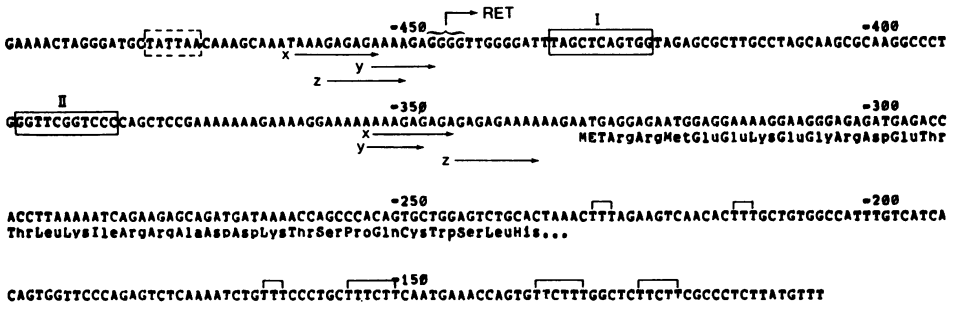


Fig. 2 Features of the Alu element within the C3(1) gene. The Figure showns 5' flanking sequences of C3(1) around the Alu repeat, the nucleotide numbers refer to distance 5' of the C3(1) $T_1$ cap site (16). The putative promoter regions for the RE are boxed, either in full for the bipartite polymerase III promoter, or dotted for the 'TATA'-like sequence. Possible RE flanking direct repeats are shown as pairs of arrows, x, y and z. The sites of transcription inititation (RET) and termination ( ) were mapped by primer extension and S1 mapping. As indicated, these transcripts could encode a small, 32 residue polypeptide (if the possibility of splicing is ignored), although it must be noted that uncapped polymerase III transcripts are not thought to be recognised by the cellular translation apparatus.

Cytoplasmic RNA samples were assayed by primer extension and S1 mapping techniques to map respectively the 5' and 3' ends of Alu transcripts. It was found that RE expression could be readily detected after gene transfer into mouse cells and some primer extension mapping experiments are shown in Fig. 3A. However, the overall results are given, with respect to the DNA sequence in Fig. 2. As can be seen, these transcripts originate at a discrete point close to the 5' end of the Alu sequence. Mapping of the 3' ends of the transcripts produced a series of S1 protected fragments whose ends mapped to clusters of 3T residues within the single copy DNA beyond the end of the repeat.

The structure of these transcripts is reminiscent of polymerase III products (5) with the 5' end close to the front segment of the intragenic promoter and with transcription apparently terminating at poly T clusters. As 3T sequences are in fact weak terminators of polymerase III transcription (29), the multiplicity of 3' ends may also be explained. However, as indicated in Fig. 2, there is a sequence 5' TATTAA 3' 30 bp upstream of the start of RE transcription which may function as a polymerase II promoter. In view of this we attempted to identify the RNA polymerase which was responsible for the observed RE expression by adding the inhibitor

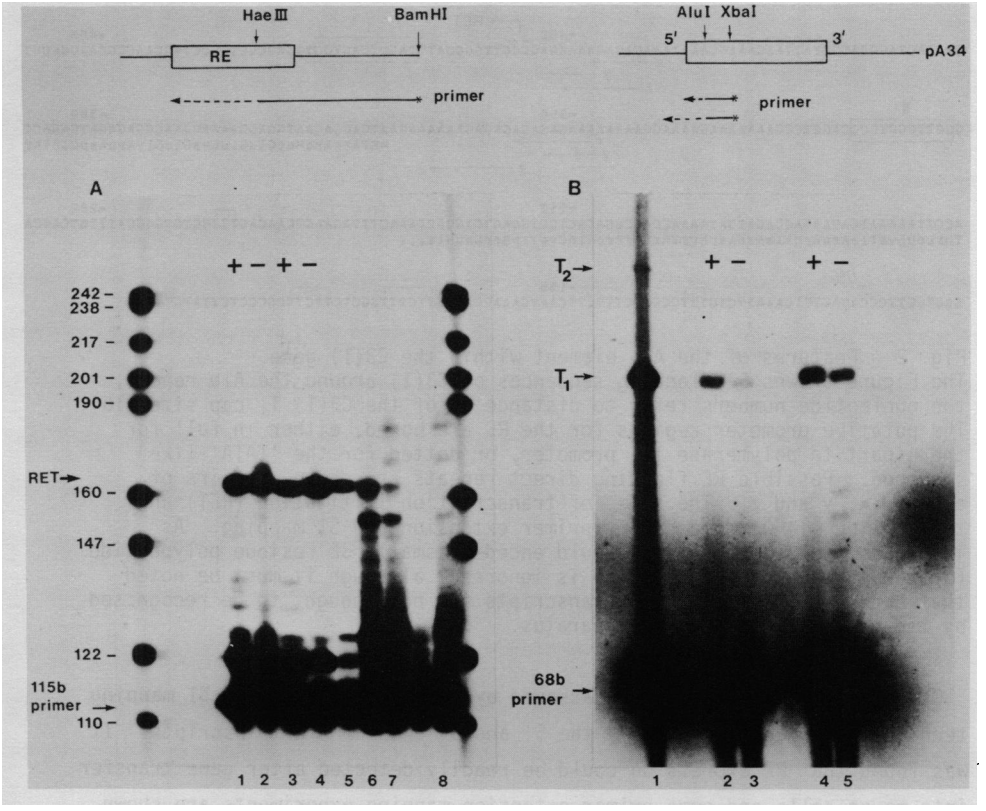


Fig. 3 Analysis of prostatic C3 and repetitive element transcripts. RNA transcripts were mapped in stably transformed cell lines or after *in vitro* transcription by using primer extension with a 115 nucleotide primer for RE (Fig. 3A) and a 68 nucleotide primer for C3(1) (Fig. 3B). The RE primer was prepared from deletion clone #5 (16) by 5' end labelling at the Bam H1 site, created at the deletion point nucleotide -292 (Fig. 2) and redigesting with Hae III at nucleotide -404. (A) Two C3(1) stably transformed cell lines, SV2 and SV6, were maintained for 4 weeks in HEPES buffered Eagle's medium with 1% steroid depleted serum alone (-) or supplemented with 3 x $10^{-8}$M testosterone (+). Cytoplasmic RNA (15 µg) from these cells was assayed by primer extension using the 115b primer at 25,000 c.p.m. per track and is shown in tracks 1 and 2 for SV2 and tracks 3 and 4 for SV6.
*In vitro* transcription mixes (100 µl) were supplemented with 40 µg/ml DNA from a C3(1) subclone treated with restriction enzymes to excise a fragment of ∿400 bp containing the Alu element. Primer extension analysis (using the 115b primer at 25,000 c.p.m. per track) of the RNA products from mixes containing o (track 6), 1 (track 7) and 200 (track 8) µg/ml α-amanitin is shown. Track 5 is a control containing 10 µg cytoplasmic RNA from L50 cells. The additional bands in tracks 7 and 8 probably represent 'end to end' transcription of the RE containing restriction fragment.

Gel bands corresponding to the 5' terminus of the Alu transcripts are marked RET.
(B) Primer extension analysis of levels of C3 transcripts within the hormonally manipulated SV2 and SV6 cells described above. All samples contained 30,000 c.p.m. of the 68b C3 primer (16) and RNA as follows: track 1, 1.0 µg prostate p(A+) RNA plus 14 µg carrier RNA; tracks 2-5 15 µg cytoplasmic RNA from SV2 (tracks 2 and 3) or SV6 (tracks 4 and 5). C3(1) transcripts initiating at $T_1$ and $T_2$ (12) are marked.

α-amanitin (30) to in vitro transcription assays. Fig. 3A (tracks 6-8) shows the results of primer extension analysis of the RNA products. The levels of correct initiations of RE transcripts were reduced at low levels of the drug, but only abolished at high concentrations. Thus, it is likely that both RNA polymerases can recognise this sequence as a transcription template, at least in vitro.

Since androgens stimulate C3 expression in vivo and after gene transfer into certain cloned S115 cells, we investigated the effect of testosterone on RE expression in responsive cells. Testosterone was without effect on RE expression (Fig. 3A tracks 1-4) in contrast to its effect on C3 expression (Fig. 3B).

Finally, we compared the expression of the Alu type III family in several rat tissues and attempted to detect transcripts from the RE associated with C3 by assaying for expression from the single copy DNA between the 3' end of the RE and the C3 cap site since this is transcribed in vitro. Although we could demonstrate Alu type III transcripts in prostate, seminal vesicle and liver we were unable to detect transcripts associated with the gene for C3 (data not shown).

## DISCUSSION

It is suggested that a third class of Alu-equivalent sequences exists in rat tissues as several examples of this element have been reported recently, both in close association with expressed genes(16,24,25) and as tissue transcripts (11). The examples of this Alu type III family show a high (> 90%) sequence homology although they originate from apparently disparate locations in the genome. This suggests that there may be some evolutionary constraint on sequence divergence, implying a functional role for these elements.

The possibility of a role for the C3 Alu element in control of the expression of the C3(1) gene has been investigated. However, although the sequence was expressed in heterologous cells, it did not respond to androgenic regulation, as was observed for C3 transcription, and there was

no evidence for in vivo expression of this particular Alu sequence. Hence, it is doubtful that Alu type III transcripts could have a functional role in C3 expression. Alternatively, the DNA sequence itself may influence local chromosome structure and therefore transcription fro the C3 gene, although this too seems unlikely. C3 expression is assumed to be coordinately regulated with the genes encoding the other components of rat prostatic steroid binding protein (namely C1 and C2) as all three gene products must come together to form a single final protein. Certainly, transcription from all three genes was influenced similarly by androgens in isolated prostate nuclei (23). However, blotting studies of cloned DNA (not shown) failed to find examples of this Alu element within the flanking DNA or introns of the C1 and C2 genes. As it is unlikely that C3 alone would be controlled by such a sequence, it seems that in the case of the C3 genes the adjacent Alu element plays no regulatory role in gene expression.

Recently these Alu type III elements have been described as identifier (ID) sequences and corresponding RNA transcripts were detected in rat brain, pituitary and peripheral nervous tissue but not in several non-neural tissues (11-13). On the basis of these studies, Sutcliffe and his colleagues have proposed that the expression of the ID sequence may be involved in tissue-specific gene activation. However, to date, there is no direct evidence that the presence or expression of particular elements, associated with specific genes, can influence their expression. In fact, its already clear that the existence of ID sequences within the genes for prolactin and growth hormone are not required for their expression in the pituitary of certain strains of rat. Within the rat prolactin gene, present at one copy per haploid genome, the insertion of a type 3 element produces an allelic variation in the 5'-flanking region (25) of the gene. As rats, which are homozygous for the absence of the element have been found, it is unlikely that Alu element is necessary for prolactin expression. Similarly, the growth hormone gene isolated from Sprague Dawley rats carries an example of the type III element within the second intron (24), but this is precisely missing from the gene isolated from Hooded rats (26).

These two examples, together with the observation of a similar element within a rat α-tubulin pseudogene (as discussed 28) provide a clear evidence that these elements may be mobile in the genome. This is perhaps reinforced by the fact that all the genomic examples of the Alu type III

element (including these sequences compared by Milner et al, 12) have short flanking direct repeats of single copy DNA. Therefore it is conceivable that the high sequence homogeneity within type III the elements resulted from their having evolved recently, and rapidly dispersed throughout the genome. The mechanism by which this occurred could well be via the return of transcripts to the genome as proposed by Jagadeeswaren et al. (8) and Van Arsdell et al. (9). Clearly these elements can produce suitable transcripts as has been shown here, and similar transcripts can be detected in some tissues in vivo (11); though, the sequence of events proposed must occur in the germ line for a new insert to persist in the population. Alternatively, the observed sequence homogeneity may be perpetuated by gene conversion events either at the genomic level or possibly between genome elements and cDNA copies of Alu transcripts. Such ideas have been proposed recently as one way of accounting for conversion-like events within the rat U3 snRNA genes and pseudogenes (31).

REFERENCES


1. Jelinek, W.R. and Schmid, C.W. (1982) Annu. Rev. Biochem. 51, 813-844
2. Fritsch, E.F., Lawn, R.M. and Maniatis, T. (1980) Cell 19, 959-972
3. Houck, C.M., Rinehart, F.P. and Schmid, C.W. (1979) J. Mol. Biol. 132, 289-306
4. Haynes, S.R. and Jelinek, W.R. (1981) Proc. Natl. Acad. Sci. USA 78, 6103-6134
5. Galli, G., Hofstetter, H. and Birnstiel, M.L. (1981) Nature 294, 626-631
6. Duncan, C.H., Jagadeeswaren, P., Wang, R.C. and Weissman, S.M. (1981) Gene 13, 183-196
7. Jelinek, W.R., Toomey, T.P., Leinward, L., Duncan, C.H., Biro, P.A., Choudary, P.V., Weissman, S.M., Rubin, C.M., Houck, C.M., Deininger, P.L. and Schmid, C.W. (1980) Proc. Natl. Acad. Sci. USA 77, 1398-1402
8. Jagadeeswaren, P., Forget, B.G. and Weissman, S.M. (1981) Cell 26, 141-142
9. Van Arsdell, S.W., Denison, R.A., Bernstein, L.B., Weiner, A.M., Manser, T. and Gesteland, R.F. (1981) Cell 26, 11-17
10. Davidson, E.H. and Britten, R.J. (1979) Science 204, 1052-1059
11. Sutcliffe, J.G., Milner, R.J., Bloom, F.E. and Lerner, R.A. (1982) Proc. Natl. Acad. Sci. USA 79, 4942-4946
12. Milner, R.J., Bloom, F.E., Lai, C., Lerner, R.A. and Sutcliffe, J.G. (1983) Proc. Natl. Acad. Sci. USA 81, 713-717
13. Sutcliffe, J.G., Milner, R.J., Gottesfeld, J.M. and Lerner, R.A. (1983) Nature 308, 237-241
14. Doolittle, W.F. and Sapienja, C. (1980) Nature 284, 601-603
15. Orgel, L.E. and Crick, F.H.C. (1980) Nature 284, 604-607
16. Hurst, H.C. and Parker, M.G. (1983) EMBO J. 2, 769-774
17. Page, M.J. and Parker, M.G. (1983) Cell 32, 495-502
18. Parker, M.G. and Mainwaring, W.I.P. (1977) Cell 12, 401-407
19. Parker, M.G., Needham, M., White, R., Hurst, H. and Page, M. (1982) Nucleic Acids Res. 10, 5121-5132

20. Sanger, F. and Coulson, A.R. (1978) FEBS Lett. 87, 107-110
21. Manley, J.L., Fire, A., Cano, A., Sharp, P.A. and Geftler, M.L. (1980) Proc. Natl. Acad. Sci. USA 77, 3855-3859
22. Jat, P., Roberts, J.W., Cowie, A. and Kamen, R. (1982) Nucleic Acids Res. 10, 871-886
23. Page, M.J. and Parker, M.G. (1982) Mol. Cell. Endocrinol. 27, 343-355
24. Barta, A., Richards, R.I., Baxter, J.D. and Shire, J. (1981) Proc. Natl. Acad. Sci. USA 78, 4867-4871
25. Schuler, L.A., Weber, J.L. and Gorski, J. (1983) Nature 305, 159-160
26. Page, G.S., Smith, S. and Goodman, H.M. (1981) Nucleic Acids Res. 9, 2084-2104
27. Sharp, P.A. (1983) Nature 301, 471-472
28. Lemischka, I. and Sharp, P.A. (1982) Nature 300, 330-335
29. Bogenhagen, D.F. and Brown, D.D. (1981) Cell 24, 261-270
30. Weinmann, R., Raskas, H.J. and Roeder, R.G. (1974) Proc. Natl. Acad. Sci. USA 71, 3426-3430
31. Bernstein, L.B., Mount, S.M. and Weiner, A.M. (1983) Cell 32, 461-472