Transcription of a stem/loop region of a tumour-amplified sequence induces bacterial aggregation

JIM HEIGHWAY* and KATE E. LEVERTON

CRC Department of Cancer Genetics, Paterson Institute for Cancer Research,
Christie Hospital and Holt Radium Institute, Manchester M20 9BX, UK

(Received 24 April 1991; revised 25 June 1991; accepted 1 July 1991)

A sequence of human DNA, amplified in a lung tumour, was shown to induce expression-dependent effects on *Escherichia coli* phenotype. Previous studies have demonstrated the induction of abnormal plasmid supercoiling and bacterial aggregation in host strains harbouring various constructs encoding this region. Subcloning identified a short stretch of 50 bp crucial for these effects. This study details further characterization of the active sequence. Computer analysis of the region predicted the formation of a stem/loop structure in transcribed RNA. Transcription, in the absence of translation, of a 22 bp sequence comprising this structure was sufficient to induce bacterial aggregation. Site-directed and random mutagenesis of the sub-region were carried out in order to identify critical factors in the induction of this phenotypic shift. It was found that changes in the loop sequence modulated activity, and mutations increasing predicted stem stability produced more active constructs. The wild-type sequence induced high level aggregation in only a limited number of strains but using the mutagenesis data, a sequence was synthesized that induced high level aggregation in most *E. coli* strains tested.

Introduction

An amplified region of human DNA was cloned from a genomic pUC19 library derived from a metastatic squamous cell lung carcinoma by virtue of weak hybridization of the sequence to a v-Ki-ras probe (Heighway & Guerts van Kessel, 1987). Expression in *E. coli* of a 1250 bp subfragment from the vector *lac* promoter was found to induce two markedly different strain-specific effects. The first was the induction of aberrant supercoiling in the host plasmid (pK42). This effect was orientation-specific with respect to the insert, and was transacting, inducing similar levels of supercoiling in a second compatible plasmid introduced into the host cell (Heighway & Guerts van Kessel, 1987). A second, more dramatic effect was noted when pK42 was introduced into the mini-cell-producing strain *E. coli* DS410 (Dougan & Sherratt, 1977). Initial growth in liquid medium was normal, but in late exponential phase the bacteria clumped, leaving a virtually clear medium with large bacterial aggregates. Regional cloning of the human DNA identified a 369 bp fragment which was capable of inducing both phenotypic alterations in *E. coli* (Heighway & Santibanez-Koref, 1989). It was assumed that a plasmid-encoded product was responsible for the effects seen, but mutation of ATG codons within a long

open reading frame encoded by this sequence and further subcloning failed to destroy clumping activity. In addition, aggregation appeared to be reduced if the region was aligned for translation. *Bal*31 deletion analysis further localized the sequence which induced aggregation to 50 bp, and restriction, filling and ligation of an internal *Sty*I site, resulting in the insertion of four bases, caused complete loss of aggregation activity. Expression constructs encoding just the wild-type 50 bp region (in contrast to pK42 encoding the entire 1250 bp sequence) showed normal coiling, suggesting that extra sequence, in addition to this region, was necessary to produce the effect. However, when the *Sty*I mutant sequence was exchanged for the wild-type in pK42, supercoiling approximated normal levels (Heighway & Santibanez-Koref, 1989). It was therefore concluded that the same sequence was important in both supercoil regulation and aggregation.

Given the results from a range of constructs it seemed very unlikely that this short sequence, important for both phenotypic changes in the host organism, encoded a peptide responsible for those effects. The most probable alternative was that the encoded mRNA molecule induced the phenotypic changes seen and the fact that translation interfered with activity suggested that the secondary structure could be important. The effects were

presumably mediated through alterations in gene expression in the host, and it seemed possible that the human sequence represented an entirely novel control mechanism. In order to define the system further, the constraints acting on the activity of the sequence were studied in more detail. Although the manner in which this RNA molecule can influence the apparently diverse processes of DNA supercoiling and bacterial aggregation is still unknown, it is possible that it also affects processes within the eukaryotic cells from which it was originally isolated.

Methods

Strains. *E. coli* strains used were as follows: JM109, JM105, HB101, BL21, DH5a, C600, CJ236, LE392 (Sambrook *et al.*, 1989) DS410 (Dougan & Sherratt, 1977), MV1190 (BioRad), SURE (Stratagene), K12 (ATCC), N4830 (Pharmacia).

Sequence analysis. Nucleic acid sequence analysis was carried out using the Genetics Computer Group Sequence Analysis Software Pack Version 6.2 (Devereux *et al.*, 1983).

Synthesis of oligonucleotides. Oligonucleotide synthesis was carried out by British Biotechnology.

Plasmid vectors. Vectors used were pUC18 and pUC19 (Yanisch-Perron *et al.*, 1985) or the pUC19 based plasmid pBS (Stratagene).

Sequencing. Plasmid DNA was prepared essentially as described by Birnboim & Doly (1979). After CsCl banding, the plasmid DNA was dialysed for 4 h in 4 changes of 1.0 l distilled water. Approx. 3 µg of each plasmid was sequenced with the Amersham Multiwell Sequencing system with the universal or reverse (BCL) primers. The double-stranded DNA was denatured by adding 0.1 vol. 2 M-NaOH, 2 mM-EDTA. After 5 min at room temperature 0.1 vol. 2 M-ammonium acetate, pH 4.8, was added, followed by 2.5 vols ethanol. After 10 min at -20 °C, the DNA was pelleted by a 5 min microcentrifuge spin. The ethanol was decanted and the pellet thoroughly air-dried. It was resuspended in 30 µl distilled water and used in the Multiwell system, as directed by the manufacturer. The only change from the provided protocol was a labelling reaction of only 3 min, at room temperature which allowed reading of the polylinker sequence.

Growth conditions. Bacteria were grown in nutrient broth (Lab M), sterilised in 500 ml batches for 1 h at 121 °C. Ampicillin and IPTG were added to 100 µg ml⁻¹ and 1 mM, respectively, and the cultures were grown with shaking at 37 °C. All comparisons were made between cultures grown in parallel using the same batch of medium.

Cloning of oligonucleotide sequences. Two cloning strategies were used. Where two complementary strands were synthesised, they were annealed by mixing 1 µg of each in 10 mM-Tris/HCl (pH 7.5), 10 mM-MgCl₂ and 10 mM-DTT. The reaction was placed in a beaker of water at 70 °C and allowed to cool to room temperature over a period of 1 h. The sequences were designed so that the annealing reaction produced cohesive termini for *Bam*HI and *Eco*RI. These DNA fragments were ligated into either pUC19 or pBS, gel-isolated after cleavage with both *Bam*HI and *Eco*RI. The recombinant molecules were transformed into competent DS410. Where the synthesised oligonucleotides were only partially complementary, the duplex was completed *in vitro*. Second strand synthesis was carried out as follows. The template (full length) strand (1 µg) was annealed to 1 µg primer strand, as above. To this was added 0.1 vol. 10× medium restriction enzyme buffer (Maniatis *et al.*,

1982), 1 µl 25 mM-dNTPs and 1 unit of Klenow fragment polymerase (Pharmacia), and the reaction was incubated for 30 min at 37 °C. Following inactivation of the polymerase by heating at 65 °C for 10 min, *Sau*3A and *Eco*RI were added, and incubation continued for 30 min at 37 °C. After heating at 65 °C for 10 min to inactivate the restriction enzymes, the DNA was ligated into either pUC19 or pBS as above. Several additional bases were added to the ends of oligonucleotides used in this way, such that the restriction enzyme recognition sites were not at the fragment termini. These were cleaved off on digestion.

PCR analysis. PCR was carried out to amplify 1230 bp fragments of human DNA corresponding to the cloned region. Primers used were:-

5' AAGCTTGAGCACTGCCAACTCAGCCT 3'

5' GGCATCCTGAATCTCTGCCTC 3'

1 µg of each primer was used to amplify 1 µg genomic DNA in a 100 µl reaction mix containing dNTPs at 250 µM and reaction buffer (IBI). After heating at 97 °C for 5 min, the *Taq* polymerase (IBI) was added and the samples cycled 30 times at 94 °C for 1 min, 65 °C for 1 min and 74 °C for 2.5 min. After amplification, 20 µl of the reaction mix was digested with *Acc*I to release a 600 bp internal fragment. Klenow polymerase and dNTPs were also added to the digest, to produce a blunt-ended molecule which was inserted into a *Sma*I-cut, phosphatased, pUC19 vector.

Results

Synthesis of a synthetic active sequence

It had been demonstrated that expression of a 50 bp fragment of the human sequence could induce bacterial aggregation in liquid culture (Heighway & Santibanez-Koref, 1989). Computer analysis of potential RNA secondary structure of this region revealed a possible 22 bp stem/loop configuration (Fig. 1). The internal *Sty*I site, which, when cut and filled, inactivated the clumping induction, lies within this putative structure. A

Fig. 1. Predicted secondary structure of RNA encoded by the human sequence.

requirement of the patent application (Heighway, 1990) was that the aggregation-inducing sequence was chemically synthesized. Towards this aim, the two complementary strands shown were produced, such that on annealing, the molecule was flanked by overhanging single-strand regions compatible with cleaved *Bam*HI and *Eco*RI sites.

5' GATCCGGAGCCCCAAGGAAAGGGGCC-
3' GCCTCGGGGTTCTTTCCCCGGG-

-TTCACCTTCCTGCTTG 3'
-AAGTGAAGGACGAAGTTAA 5'

The double-stranded molecule was then ligated into *Bam*HI/*Eco*RI-cleaved pBS vector and used to transform DS410. Thirty ampicillin-resistant colonies, screened for aggregation in liquid culture, showed, with two exceptions, normal growth. Recombinant plasmids were sequenced to confirm fidelity of the insert. These results confirmed that if the construction was as intended, normal growth resulted. This was not surprising, as the insert was aligned for translation, and previous work suggested that this interfered with the induction of aggregation (Heighway & Santibanez-Koref, 1989). The two constructs (pSA31, pSA51) that did induce aggregation were shown to be the result of aberrant ligation/recombination events. Sequencing of pSA31 revealed that a deletion event had occurred so that the 5' end of the oligonucleotide sequence was ligated directly to a position several bases downstream of the point of initiation of transcription of *lacZ*. The second construct inducing aggregation (pSA51) was shown to be a tandem repeat of the oligonucleotide resulting from a spurious ligation event between the *Eco*RI end of one sequence and the *Bam*HI end of another, presumably occurring *in vivo* after transformation of the host with a linear construct.

In order to test directly whether the normal phenotype of cells carrying the 43 bp fragment was dependent on translation of the product mRNA, an in-frame translational stop codon was introduced between the *lacZ* initiation codon and the inserted sequence. To achieve this, a construct with the intended sequence (pSA6) was cut at the upstream *Hind*III site. The ends were filled, and the linear plasmid re-ligated and transformed into DS410. Sequencing confirmed the presence of the in-frame stop codon and this construct (pK710) induced a high level of aggregation in liquid culture.

Directed mutagenesis of the basic sequence

Computer analysis predicted that RNA encoded by the human sequence would form a stem loop structure (Fig 1). It was decided to remove the bulging A from the

putative stem region to see whether this resulted in a change in clumping induction. Two partially complementary overlapping strands were synthesized such that the desired mutation was present in a single-stranded region of the annealed duplex.

5' ATGGATCCGGGCCCAAGGAAAGGGGCC-

-CTTCACTTCCTGCTTG 3'
3' GAAGTGAAGGACGAAGTTAAGAAA 5'

The full double-stranded molecule was then synthesized and cleaved with *Sau*3A/*Eco*RI to generate a product directly comparable to the pSA6 insert. This strategy of a template and primer strand meant that only one strand needed to be synthesized for each new sequence, the complement being extended from the primer. It was found that deletion of the bulging A from the stem region increased the power of the recombinant plasmid (pK441) to induce aggregation of DS410. When pK441 was introduced into the host, the construct was able to induce aggregation when aligned for translation, in contrast to the wild-type sequence. However, the aggregation seen was at a low level in that the clumps were very small and the medium not totally clear. As in pK710, activity was increased following the introduction of an in-frame stop codon to block translation.

The potential stem was elongated in a further construct by the addition of three Cs and Gs to the base of the structure.

5' ATGGATCCCGGGGCCCAAGGAAAGGGG-
-CCCGGGTTCACCTTCCTGCTTG 3'

This further increased the activity such that high level aggregation (large clumps and clear medium) was induced even though the sequence was aligned for translation. In order to determine whether the primary stem structure was important an oligonucleotide was constructed in which C residues were exchanged for Gs and vice versa.

5' ATGGATCCGGACGGGGAAGGAAACCCCG-
-CCTTCACTTCCTGCTTG 3'

Although the predicted secondary structure and stability were similar to the wild-type, the new construct was only able to induce a very low level of aggregation, which manifested as a slight granular appearance in the culture medium. This suggested that at least part of the primary structure of the stem was important for full activity.

To confirm that expression of the stem/loop region alone was sufficient to induce clumping, two complementary sequences were annealed such that they encoded this structure, flanked by *Xba*I/*Eco*RI sites.

Fig. 2. Random mutagenesis protocol. The ends of the template oligonucleotide were synthesised using pure dNTP solutions. The centre section (as indicated) was made using mixed dNTP stocks, such that there was only an 85% chance of incorporation of the correct base. Complimentary strand synthesis and cloning were carried out as described in Methods.

Fig. 3. Results of random mutagenesis. Sequences in (a) increase activity over wild-type levels (top line, bold) while those in (b) abolish it. Substitutions are in lower case, deletions represented by a dash and insertions underlined.

When inserted into a pBS vector (with the *Hind*III site filled as above) this new construct, encoding just 22 bp of the human sequence, was found to induce aggregation to the same level as pK710.

Random mutagenesis

The synthetic sequence, when inserted in-frame with the *lac* ATG (pSA6), did not induce aggregation in DS410. However, if the stem mismatch was removed (p441), activity was detected in a comparable construct. It was reasoned that the system could therefore be used in random mutagenesis experiments in order to identify mutations that increased or decreased the ability of the sequence to induce aggregation. An oligonucleotide was synthesised such that the ends were constant but the centre (stem/loop) region was variable. This guarantees a constant primer-binding segment to allow second strand synthesis, and constant restriction enzyme sites at the ends (Fig. 2). In the centre of the molecule, the synthesis was carried out using nucleotide mixes spiked with 5% of each of the other three dNTPs. This figure was chosen to

give a reasonable frequency of mutant sequences in the ligation mix. After annealing, extension and digestion, the reaction mix was split and ligated into either pBS (ligation a) or pBS with a filled *Hind*III site (ligation b). After transformation of the two ligation mixes into DS410, it was therefore possible to screen for mutations that enhanced or decreased the activity of the sequence. In wild-type form, the construct ligated into pBS would not induce aggregation, in contrast to the same sequence in the *Hind*III-filled vector where aggregation induction would be expected. Individual transformants (approx. 400) were screened from both ligations for clumping induction, and constructs not showing the wild-type activity were initially analysed by gel electrophoresis to confirm the insert and approximate size, and subsequently by sequence analysis.

The constructs with enhanced clumping activity were found to be mainly tandem repeats of the oligonucleotide sequence (similar to pSA51) and so provided no new information. However a number of monomer inserts that induced aggregation were identified (Fig. 3a). The predominant and most effective mutation was deletion of the final A of the putative loop. Alteration of this base to C also enhanced activity. The deletion of this A produced a construct that was able to induce aggregation to a similar level to the original plasmid pK42 (encoding

Transformants from ligation *b* showing no aggregation in culture, were analysed in the same way. Some showed obvious deletion of the vector or absence of insert. Of the rest, sequence analysis showed that most had either multiple base changes or deletions of small regions of the putative stem/loop, several of which are detailed in Fig. 3*b*. In particular, one inactive sequence had a single base change within the putative stem (G to T) which would theoretically result in a less stable structure. Also inactivating were insertions of single A residues to the loop and a G-T point mutation

Using the information from the mutagenesis experiments, a sequence was synthesized that incorporated alterations predicted to increase activity. The final loop A was omitted, and the stem region was extended to 16 GC pairs. Two strands were synthesized and annealed to generate the duplex below:

-GGGGCCCCCCCCCG 3'
-CCCCGGGGGGGGGGCTTAA 5'

Eukaryotic sequences with activity

RNA. These were bases 743–764 of the cDNA encoding the sweet potato phenylalanine ammonia-lyase gene (Tanaka *et al.*, 1989) and bases 1288–1307 of the rat GABA receptor gene (Ymer *et al.*, 1989). A synthetic molecule was constructed, encoding the plant sequence such that this region was flanked by *Bam*HI/*Eco*RI sites and this was inserted into pBS.

-TTGCCCTCG 3'
-AACGGGAGCTTAA 5'

5' GATCCATGCCCAAGGAAGGGCATG 3'
3' GTACGGGTTCCTTCCCGTACCTAG 5'

PCR analysis

In order to confirm that the human sequence isolated had not been modified during tumorigenesis, PCR was carried out using peripheral blood DNA from five individuals. A 1230 bp fragment was amplified and cleaved with *AccI* to release a 600 bp internal sequence including the active region. The fragment was made blunt-ended and cloned into pUC19. After transformation into DS410, clones were analysed for clumping induction. Constructs from all five individuals induced aggregation and analysis verified the expected orientation and sequence, demonstrating that the stem/loop region was a normal component of the genotype.

Discussion

The aim of this study was to define further the 50 bp region previously shown to be crucial to the production of aberrant plasmid supercoiling and sufficient to induce bacterial aggregation in *E. coli* (Heighway & Guerts van Kessel, 1987; Heighway & Santibanez-Koref, 1989). Computer analysis of the sequence indicated that the RNA was most likely to adopt a 22 base stem/loop structure within the transcribed message. It had previously been shown that cutting, filling and re-ligating the *StyI* site within the 50 bp region inactivated the clumping induction and produced near-normally supercoiled plasmid. This site was within the putative structure and the filling reaction would cause the addition of four bases to the loop. Further evidence for the importance of the stem/loop came from earlier *Bal31* deletion studies (Heighway & Santibanez-Koref, 1989). Several plasmids were produced that induced a reduced level of aggregation compared to controls. These were shown to involve deletions removing either one or two bases from the downstream stem section which would result in a destabilization of the structure. Later work with the plant and rat sequences showed that the primary structure at the base of the stem was not critical. The fact that deletion of stem bases (effective shortening of the stem) reduces activity is further evidence that this structure forms and is important for function.

The chemical synthesis and cloning of this region confirmed that if the active sequence was aligned for translation, liquid growth was normal with no clumping induced. The two spurious clones pSA31 and pSA51 added more information. The fact that pSA31 induced aggregation effectively removed all possibility that an encoded peptide was responsible for the phenotype. The result with pSA51 showed that a tandem repeat of the active region increased activity. This could have been due to inhibition of 3' to 5' degradation of the RNA when the host RNAase encountered the first stem/loop or simply that each molecule contained two effector regions. Producing constructs with greater numbers of tandem repeats may therefore further increase effectiveness.

The mutagenesis experiments produced some unexpected results. The protocol was designed to select variants with point mutations, but most of the clones with aberrant activity were found to be the result of deletions or insertions of bases. This was presumably due to errors in the synthesis reaction of the oligonucleotide template. Taken together, the mutation data strongly suggest that the ability to form stable stem/loop structures is essential for activity, as is the sequence of the loop and perhaps a portion of the stem. It seems likely

that translation of the RNA interferes with this structure in some way, preventing its action. Using this information, a sequence was derived that induced aggregation in the majority of the *E. coli* strains tested. Notable exceptions were HB101 and DH5a. There is no obvious factor in the genotypes of these strains that might explain the lack of response to the cloned region and it is hoped to use them to identify bacterial genes involved in induction of the two effects.

The aggregation of *E. coli* would seem to be a unique response, possibly involving a novel mechanism, and the effect on pathogenicity has yet to be determined. The supercoil abnormalities induced, however, are similar to effects seen in starved organisms (Balke & Gralla, 1987). The mechanism by which the phenotypic changes are induced is not known. Preliminary studies have indicated no apparent effect on global gene expression and the most likely explanation of the activity is that specific changes are made in expression resulting in altered levels of particular gene products. It is worth noting that the natural antisense sequences have stem/loop regions critical to activity (Simons, 1988). However the lack of requirement for base specificity of at least half of the stem might suggest either that the target sequence is very small or that a novel mechanism is involved. A further possibility is that the sequence binds a non-nucleic-acid target molecule within the cell. If this is the case it may be feasible to co-isolate the molecule with the transcribed RNA. It is hoped that by using transposon mediated mutagenesis and inverse PCR it will be possible to isolate target genes which should lead to a better understanding of the mechanism involved. If, as expected, the process involves the perturbation of gene expression this will be of great interest, given the origin of the sequence and its dramatic effect on phenotype in *E. coli*. The question of potential activity in human systems would then be very important.

The computer search using the optimum core sequence yielded 16 exact matches. In two of these there was also present an extended flanking region capable of forming a stem/loop structure in RNA. Given that in a random system the probability of having matching bases is one in four for each stem residue (five pairs in the sweet potato and four pairs in the GABA sequence) the probability of these sequences having arisen by chance is very low. This may indicate a functional constraint on activity and cellular role for these structures in eukaryotic cells.

The aggregation of *E. coli* from liquid culture has commercial implications (Heighway, 1990). PCR analysis confirmed that the region was not a tumour-specific sequence, which might have precluded its use. However, the implications of amplification of the sequence in tumorigenesis remain to be determined.

We would like to thank Mr N. Barron for the numerous plasmid preparations that he endured and the Cancer Research Campaign and CRC Technology for funding the work.

References

- BALKE, V. L. & GRALLA, J. D. (1987). Changes in the linking number of supercoiled DNA accompany growth transitions in *Escherichia coli*. *Journal of Bacteriology* **169**, 4499–4506.
- BIRNBOIM, H. C. & DOLY, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research* **7**, 1513–1525.
- DEVEREUX, J., HAEBERLI, P. & SMITHIES, O. (1983). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387–395.
- DOUGAN, G. & SHERRATT, D. (1977). The transposon TnI as a probe for ColE1 structure and function. *Molecular and General Genetics* **151**, 151–160.
- HEIGHWAY, J. (1990). Genetically engineered processing of micro-organism cultures and DNA related thereto. UK Patent Application no. 9004896.8.
- HEIGHWAY, J. & GEURTS VAN KESSEL, A. H. M. (1987). Isolation of a human genomic fragment, co-amplified with c-Ki-ras, that affects plasmid DNA supercoiling in *E. coli*. *Nucleic Acids Research* **15**, 3411–3420.
- HEIGHWAY, J. & SANTIBANEZ-KOREF, M. F. (1989). Unexpected effects on bacterial phenotype induced by expression of a tumour amplified sequence. *Nucleic Acids Research* **17**, 6893–6901.
- MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning: a Laboratory Manual*, p. 104. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd Edn, pp. 9–13. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- SIMONS, R. W. (1988). Naturally occurring antisense RNA control- a brief review. *Gene* **72**, 35–44.
- TANAKA, Y., MATSUOKA, M., YAMANO, N., OHASHI, Y., KANO-MURAKAMI, Y. & OZEKI, Y. (1989). Structure and characterisation of a cDNA clone for phenylalanine ammonia-lyase from cut injured roots of sweet potato. *Plant Physiology* **90**, 1403–1407.
- YANISCH-PERRON, C., VIEIRA, J. & MESSING, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.
- YMER, S., SCHOFIELD, P. R., DRAGUHN, A., WERNER, P., KOHLER, M. & SEEBURG, P. H. (1989). GABA receptor B subunit heterogeneity: functional expression of cloned cDNAs. *EMBO Journal* **8**, 1665–1670.