Volume 15 Number 3 1987 Nucleic Acids Research

Shufflon: multi-inversion of four contiguous DNA segments of plasmid R64 creates seven different open reading frames

Teruya Komano, Akihiro Kubo* and Taizo Nisioka

Department of Biology, Tokyo Metropolitan University, Fukazawa, Setagaya-ku, Tokyo 158, Japan

Received November 17, 1986; Accepted January 8, 1987

ABSTRACT

The IncIα plasmid R64 was found to bear a highly mobile DNA segment which was designated as a clustered inversion region (J. Bacteriol. 165, 94-100, 1986). The clustered inversion region consists of four DNA segments designated respectively as A, B, C and D which differ in molecular size and restriction sites. The four DNA segments invert independently or in groups resulting in a complex DNA rearrangement. We now show the nucleotide sequence of the clustered inversion region of R64. The present results suggest that the clustered inversion region is a biological switch to select one of seven open reading frames whose primary structures at the region proximal to N-termini are constant while those at the C-terminal region are variable. A name, "Shufflon" was proposed to call this kind of the clustered inversion region.

INTRODUCTION

In a number of procaryotic systems, the inversion of a specific DNA segment has been shown to regulate gene expression (for reviews, see references 1 and 2). In Salmonella typhimurium, the inversion of H segment switches the alternative production of H1 or H2 flagellin (3, 4). The G and C inversions control the host range of phages Mu and P1, respectively (5-8). Similar P inversion was found on the Escherichia coli chromosome (9). These inversions are well characterized and are shown to be accomplished by site-specific recombination. They are closely related since the sequences of DNA crossover sites are homologous (4, 7-9) and the genes of site-specific recombinases (hin, gin, cin, and pin) are also homologous and complement each other (10). Production of type 1 fimbriae of E. coli is also regulated by an inversion of a 314-bp DNA segment (11).

In our previous works (12, 13), we have found and cloned a highly mobile DNA segment in IncIα plamid R64 (pKK009). Restriction enzyme analysis showed that pKK009 DNA was a mixture of six or more DNA species. When a part of pKK009 DNA was deleted, 33 different types of fixed plasmids (pKK010-series plasmids) were obtained among 58 clones tested. Detailed comparison of the

restriction maps of pKK010-series plasmids led us to propose a model in which four DNA segments, designated as A, B, C and D, invert independently or in groups, so that the arrangement of the four segments change randomly. It was shown that the DNA rearrangement was mediated by a gene fuction, designated as rci, which was located near the inversion region. This DNA rearrangement was designated as a clustered inversion region.

In this paper, we have determined the nucleotide sequence of the clustered inversion region of R64. There are seven 19-bp repeats which separate the four DNA segments. The results indicate that the DNA rearrangement is accomplished by a series of site-specific recombinations between any two repeats which lie in the opposite direction. On the basis of the analysis of open reading frames of this region, we have proposed a model in which the clustered inversion region is a biological switch to select one of the seven proteins whose primary structure at the region proximal to N-termini are constant while those at the C-terminal region are variable. A name, "Shufflon" was proposed to call this kind of the clustered inversion.

MATERIALS AND METHODS

Bacteria, phages and plasmids

Escherichia coli JM83 *ara* Δ(*lac-proAB*) *rpsL* ϕ80 d*lacZ*Δ*M15* and JM103 (*lac-proAB*) *supE thi rpsL sbcB15 endA hspR4* F' *traD36 proAB lacI*^{q} *Z*Δ*M15* (14) were used. Sequencing vectors M13mp18 and M13mp19 (14) were used. Plasmid pKK010-85 (13) was used for DNA sequence determination.

DNA sequence determination

The DNA sequence was determined by the method of Messing (15).

Materials

Restriction enzymes and M13 sequencing kit were obtained from Takara Shuzo, Boehringer Mannheim and Bethesda Research laboratories. [α-^{32}P]dCTP (400 Ci/mmol) was from Amersham Corp.

RESULTS AND DISCUSSION

DNA sequence of the clustered inversion region in R64

Restriction map of the inversion region of a fixed plasmid, pKK010-58, which has arrangement of DNA segments ADCb was constructed as shown in Figure 1. The nucleotide sequence of the inversion region of pKK010-85 was determined according to the strategy shown in Figure 1. Figure 2 shows a sequence of 3421 bp from the HpaI site to the third EcoRV site. Every nuclotide was determined in both directions using overlapping fragments.

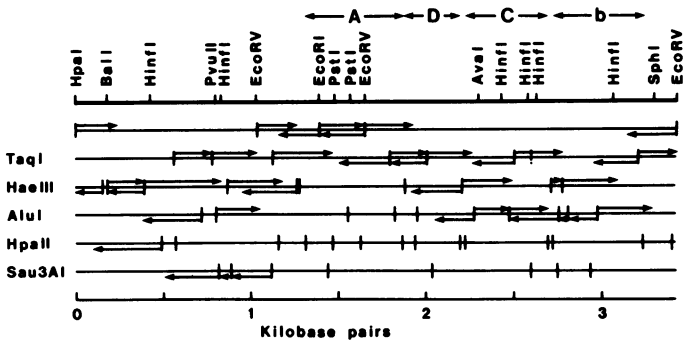

Figure 1. Restriction map and sequencing strategy. The restriction map of pKK010-85 (13) is shown by a bold line. Below this, the sequencing strategy is shown. At the top, the names of the four DNA segments are indicated. DNA segments in opposite direction are indicated by small letters.

There are seven 19-bp repeats oriented in the same or opposite direction as shown by the large arrows in Figure 2. They flank each of the four DNA segments: For example, segment A is flanked by repeats 1 and 2. Likewise, segments D, C and B are flanked by repeats 2 and 3, 4 and 5, and 6 and 7, respectively. DNA rearrangement might occur through a series of site-specific homologous recombinations between any two repeats which lie in the opposite direction. This prediction was supported by the analyses of the nuclotide sequence at some different junctions of the four segments in the other pKK010-series plasmids (S. Kim & T. Komano, unpublished results). Segments A, B and C are bracketed by inverted repeats whereas segment D is bracketed by direct repeats. This fact is consistent with the previous observation that segment D takes only one orientation when it lies at the leftmost or the rightmost positions of the inversion region (13). Homologous recombination seems to take place between both polarities of the inverted repeats (→---------← and ←---------→), since segment D takes both orientation when it lies at the 2nd or 3rd positions of the inversion region (13). However, recombination is not likely to occur between direct repeats, since deletion of DNA segments was not detected from the inverting plasmid pKK009 (13). The DNA sequences of 8-bp regions between repeats 3 and 4, and those between 5 and 6 are completely identical, though their polarity is reversed.

DNA crossover sites

Figure 3 compares the seven 19-bp repeat sequences. Left-hand 13-bp and right-hand 3-bp sequences are completely conserved, whereas the consensus AC

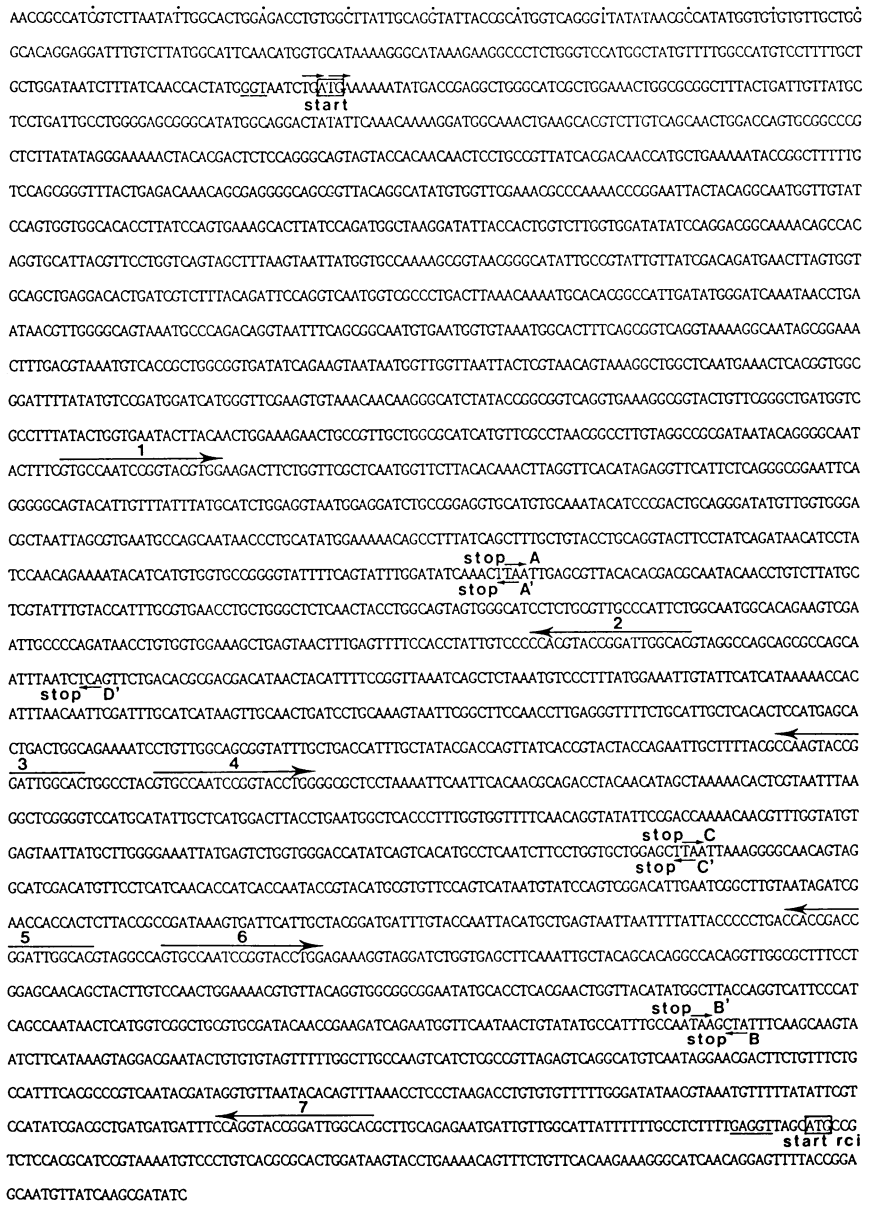

Figure 2. Nucleotide sequence of the clustered inversion region of R64. HpaI-EcoRV region of pKK010-85 was sequenced by the M13 dideoxy chain termination method. Seven 19-bp repeat sequences are indicated by large arrows. Initiation codon of ORF are boxed. Putative Shine-Dalgarno sequences are underlined. Termination codon are indicated by small arrows. In addition to the seven 19-bp repeats, 17-bp direct repeats (position 973-989 and 1162-1178) with 1-bp mismatch are detectable in this region.

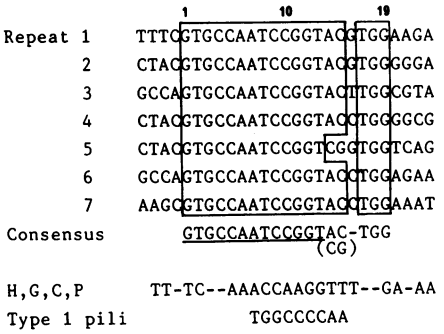

Figure 3. Alignment of repeat sequences. The seven 19-bp repeat sequences are aligned in the same direction. Conserved regions are boxed. Consensus sequence of the repeats is shown together with the crossover site of H, G, C and P inversion regions (4, 7-9) and that of type 1 fimbriae inversion region (11). Crossover of the recombination in the clustered inversion region is supposed to occur somewhere within the underlined 13-bp in consensus sequence (see text).

sequence at position 14 and 15 is replaced by CG in repeat 5 and the 16th base is C, G or T. These sequences show no remarkable homology with the DNA crossover sites of H, G, C and P regions or that of the control region of the E. coli type 1 fimbriae gene. Repeats 4, 6 and 7 have a KpnI site. Hence, we mapped KpnI sites in the inversion region of all 58 pKK010-series plasmids described in the previous paper (13). In all cases, KpnI sites are located at both ends of segment B and one particular end of segment C (data not shown), indicating that the actual crossover site for the recombination event is somewhere within the conserved 13-bp sequence. The structural difference in the repeat sequences seems to be the reason for the unequal frequencies of the inversion of the four segments as observed previously (13). For example, the frequencies of plasmids having the A segment at the leftmost position of the inversion region was about two thirds of the population at equilibrium.

Analysis of open reading frames

Open reading frame (ORF) of this region was analyzed (Figures 2, 4). A long ORF preceded by a possible Shine-Dalgarno sequence (GGT) starts at position 237, proceeds rightward beyond repeat 1, and stops at position 1658 which lies inside segment A. Segment A has another ORF extending from repeat 2 to position 1661. If segment A inverts between repeats 1 and 2, the ORF after repeat 2 connects in frame with the ORF starting from position 237 (Figure 4-II). Also, segments B and C each have two inward ORFs from the repeats at both ends, while segment D has only one ORF. When these ORFs move to the leftmost position of the inversion region, they connect in frame with

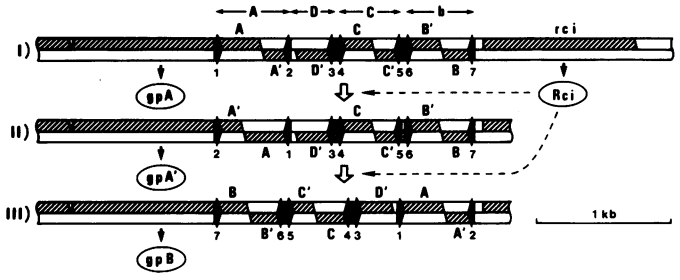

Figure 4. Model for switching of the seven genes. I) represents gene organization of pKK010-85. Black arrows represent the seven 19-bp repeats. The four DNA segments are indicated above the figure. ORFs in the left-to-right or right-to-left direction are indicated by top or bottom hatched area of the figure, respectively. Inversion of segment A, mediated by the rci product Rci, results in plasmid II. Subsequent inversion of whole segments results in plasmid III. Random inversion of the four segments independently or in groups results in a mixture of 1152 isomers ($= 4! \times 2^5 \times 1.5$) if one considers even the direction of small 8-bp segments between repeats 3 and 4 and between 5 and 6.

the ORF starting from postion 237 (one example is shown in Figure 4-III). These results suggest that the clustered inversion region has a function to select one of seven genes, giving rise to produce one of seven proteins of which N-terminal regions are constant, while C-terminal regions are variable. It is noteworthy that stop codons in ORFs from both repeats in the segments A, B or C are overlapping or very close with repect to their sequence (Figure 2). We could not find a promoter structure for the ORF starting from position 237. Instead, there is another ORF continuing far from position 1, which stops with two consecutive nonsenese codons overlapping with the initiation codon of the long ORF. The ORF starting from position 3295 appears to be the coding region of rci, a gene for a site-specific recombinase. Preliminary results of sequencing of this region indicate the presence of an ORF of 384 codons (A. Kusukawa & T. Komano, unpublished results).

Figure 4 shows a model for switching of the seven genes by inversions. Inversion between repeats 1 and 2 of plasmid I in which gene A is expressed, results in plasmid II in which the gene is changed to A'. Further inversion of plasmid II between repeats 2 and 7 results in plasmid III in which gene B is expressed. Random inversions between any two of the inverted repeats result in expression of each of the seven genes. We would like to propose that this kind of the clustered inversion might be called a "Shufflon".

Deduced amino acid sequences of the constant and variable regions are

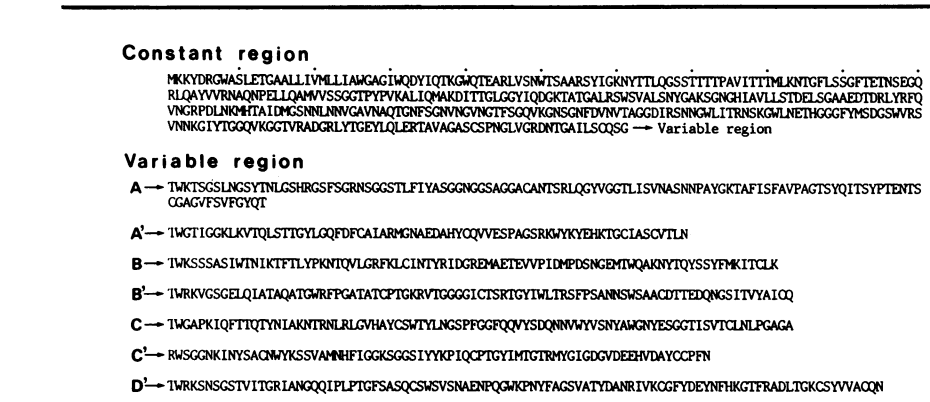

Figure 5. Deduced amino acid sequences of the product coded for by the variable genes. Amino acid sequences of constant region and variable region are indicated.

shown in Figure 5. Constant regions consist of 361 amino acids, whereas the number of residues in the variable regions fluctuate between 69 and 113. No remarkable homology is noticed among seven variable regions. The amino terminal region of a constant region consists of a basic amino acid part, a hydrophobic amino acid domain and AlaTrpGlyAlaGly (cleavage site ?) in succession. This structure somewhat resembles the signal peptide of secretory proteins except for the occurrence of aspartic acid and glutamic acid (16). However, it is too premature to conclude that these genes encode secretory proteins.

At present we do not know the biological significance of this DNA region. The wide distribution of similar DNA regions around the center of the "core" region of many IncIα plasmids (T. Komano, manuscript in preparation) suggests an important role of this region in the basic function of the IncIα plasmids. Further investigation is in progress for the elucidation of this matter.

ACKNOWLEDGEMENTS

We are grateful to Dr. Masayori Inouye for his valuable discussion. We thank Dr. Shiro Tomino and Dr. Bert Lampson for their critical reading of the manuscript and Dr. Teiichi Fruichi for computer analysis.

* Present address: Environmental Biology Division, the National Institute for Environmental Studies, Yatabe-machi, Tsukuba, Ibaraki 305, Japan

REFERENCES

1. Silverman, M. and Simon, M. (1983) In Shapiro, J. A. (ed), Mobile Genetic Elements, Academic Press, New York, pp. 537-557.
2. Plasterk, R. H. A. and van de Putte, P. (1984) Biochim. Biophys. Acta 782, 111-119.
3. Zieg, J., Silverman, M., Hilmen, M. and Simon, M. (1977) Science 196, 170-172.
4. Johnson, R. C. and Simon, M. I. (1985) Cell 41, 781-791.
5. Kamp, D., Kahmann, R., Zipser, D., Broker, T. R. and Chow, L. T. (1978) Nature 271, 577-580.
6. Hiestand-Nauer, R. and Iida, S. (1983) EMBO J. 2, 1733-1740.
7. Kahmann, R., Rudt, F., Koch, C. and Mertens, G. (1985) Cell 41, 771-780.
8. Iida, S. and Hiestand-Nauer, R. (1986) Cell 45, 71-79.
9. Plasterk, R. H. A. and van de Putte, P. (1985) EMBO J. 4, 237-242.
10. Kutsukake, K. and Iino, T. (1980) Proc. Natl. Acad. Sci. USA 77, 7338-7341.
11. Abraham, J. M., Freitag, C. S., Clements, J. R. and Eisenstein, B. I. (1985) Proc. Natl. Acad. Sci. USA 82, 5724-5727.
12. Furuichi, T., Komano, T. and Nisioka, T. (1984) J. Bacteriol. 158, 997-1004.
13. Komano, T., Kubo, A., Kayanuma, T., Furuichi, T. and Nisioka, T. (1986) J. Bacteriol. 165, 94-100.
14. Yanisch-Perron, C., Viera, J. and Messing, J. (1985) Gene 33, 103-119.
15. Messing, J. (1983) Methods Enzymol. 101, 20-78.
16. Duffaud, G. D., Lehnhardt, S. K., March, P. E. and Inouye, M. (1985) In Current Topics in Membranes and Transport, Academic Press, New York, Vol. 24, pp. 65-103.