The EMBO Journal vol.3 no.6 pp.1417 - 1421, 1984

A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*

Eric Gilson, Jean-Marie Clément[2], Douglas Brutlag[1] and Maurice Hofnung*

Unité de Programmation Moléculaire et Toxicologie Génétique, CNRS LA271, INSERM U163, Institut Pasteur, 28, Rue du Docteur Roux, 75015 Paris, France, and [1]Department of Biochemistry, Stanford University Medical Center, Stanford, CA 94305, USA

[2]Present address: Salk Institute for Biological Studies, La Jolla, CA 92138, USA

*To whom reprint requests should be sent

Communicated by M.Hofnung

We report the properties of 67 members of a family of dispersed repetitive palindromic extragenic bacterial DNA sequences. These sequences, called palindromic units, appear to be present at least several hundred times outside structural genes on the *Escherichia coli* chromosome. They are found either in clusters – as in a previously described intercistronic element – or in single occurrences. They are not only found within an operon but also between different operons, including between convergent ones. The palindromic units could yield a stem and loop structure at the level of DNA or RNA. The base of the stem is made of eight remarkably conserved base pairs while the rest varies somewhat in length and sequence. We analyse the data available on the palindromic units and we speculate on their possible roles with emphasis on transcription and mRNA stability or processing, as well as on their possible relation to transposition elements and the modular evolution of the genome.

Key words: palindromic units/*E. coli*/intercistronic regions/DNA

Introduction

We described recently a genetic element common to several intercistronic regions of bacterial operons (Higgins *et al.*, 1982a; Clément and Hofnung, 1981). This element was composed of a succession of up to three occurrences of a palindromic unit (p.u.). Most of the occurrences could correspond to a potentially stable stem and loop DNA or RNA structure. The lower part of the stem is composed of eight highly conserved base pairs (G-T is recorded as a pair because of the potential G-U pairing) and a consensus sequence – which we call here the consensus stem – could be derived from it. This highly conserved part of the stem is generally terminated by a C-T mismatch at the end nearest the hairpin loop. We made a systematic investigation of a series of DNA sequences and we discovered that the p.u. is not only found in repeats between genes of the same operon but also in single occurrences and between different operons. By analyzing the available data in the literature we substantiate a possible involvement in gene expression. We also discuss how the high number of occurrences of the p.u.s, their localization outside structural genes and the homology to the ends of transposons, presented here, are compatible with a possible role in genome evolution.

Results and Discussion

Location and structure

We made a computer search for the consensus stem of the p.u. as well as for sequences differing from the consensus stem by at most four bases (for details see Materials and methods). We also screened, by eye, non-coding regions of bacterial sequences not contained in the data base or adjacent to the occurrences detected with the computer. Figure 1 shows the positions of the occurrences with respect to known genes. Figure 2 lists their DNA sequences. References are given in the legend to Figure 1. We found 10 triple occurrences, 10 double occurrences, and one occurrence (*his*G) where the p.u. is present one and half times. In these 51 cases p.u.s are exclusively located in non-coding regions of bacterial DNA. We found 33 single occurrences in bacterial and phage DNA; about half of them are in non-coding regions (Table I). In eukaryotic DNA, we found only single occurrences; 20 in all.

We found 23 occurrences identical to the consensus stem (abbreviated as 'consensus occurrences'). This represents a remarkable conservation, much higher than that of *E. coli* promoters for example. The upper part of the stem is less constant but can often be extended up to five less conserved base pairs. A variable loop, generally smaller than five nucleotides, is rather AT rich while the stem is GC rich. Successive repeat of the p.u. often present the sequence CTACG/A at their 3′ end which permits us to orient the unit concurrently with the C-T mismatch within the stem. CTACG/A is mostly found in multiple occurrences of the p.u. and may therefore play a role in the generation of the repeat. The consensus sequence includes either T or G in the fourth position of the stem pairing with its exact complement A or C. T and G almost always alternate in successive occurrences at this position (the only exceptions are the *rpl*L-*rpo*B and $tRNA^{Gln2}$ clusters).

These 23 consensus occurrences correspond exclusively to bacterial extragenic sequences: 22 from *E. coli* and one (*his*G) from *Salmonella*. Since we screened ~3–5% of the *E. coli* chromosome we can rule out a fortuitous event and, using a rough extrapolation, predict that the consensus stem should be present at least several hundred times on the *E. coli* chromosome. Such a minimal figure is compatible with the fact that in some cases (*mal*B, *ndh* and *met*JBLF) several clusters of p.u.s are found in the same genetic system.

The number of occurrences presenting one difference from the consensus stem was 13 in prokaryotic DNA. All these occurrences are extragenic, like the consensus occurrence. In eukaryotic DNA, three occurrences differing at one position from the consensus stem were found in ~1.5 kilobases; this number is within statistical expectation for a random event.

Occurrences with two or more differences from the consensus stem are expected to be quite frequent on a purely statistical basis. It is thus not surprising that they are intragenic as

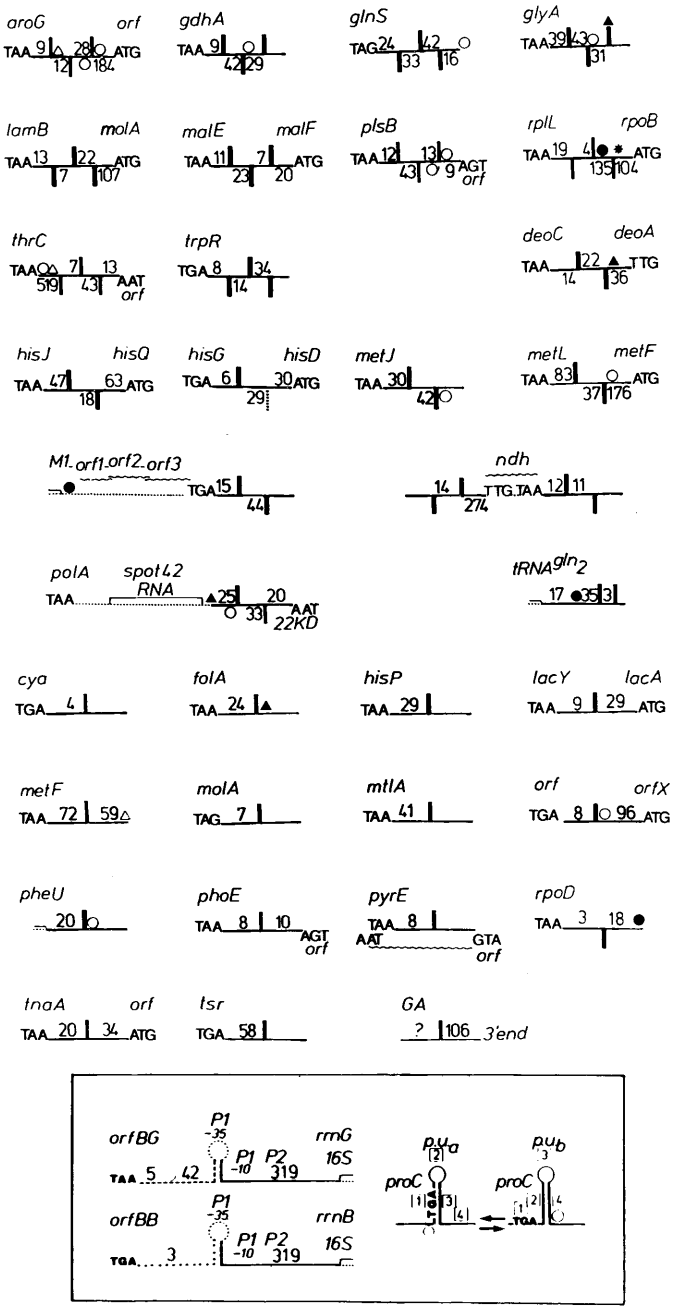

Fig. 1. Positions of bacterial extragenic palindromic units. The DNA sequences around p.u.s are indicated by continuous horizontal lines. DNA flanking regions are drawn as dotted lines. Each p.u. is indicated with a continuous vertical bar. In the case of *his*G the dotted vertical bar represents one branch of the stem. The bar is pointed upwards when the sequence given on Figure 2 is on the putative coding strand (named a, b, c... on Figure 2). It is pointing downwards in the other case (named a', b', c'...). Whenever possible, genes on both sides of p.u.s are indicated, as well as the relevant translation initiation and stop codons. P.u.s are named according to the gene on the left, usually but not always upstream in transcription and ordered from left to right (see Figure 2 and legend). The two p.u.s upstream of *ndh* are called *ndh*-up and the one upstream of *orf*X is called *orf*X-up. Genes transcribed from left to right are named above the horizontal lines; genes transcribed in the opposite orientation are named under the lines. The nomenclatures for genes are taken from the original publications. Numbers designate the lengths in base pairs starting from the bottom of the stems between p.u.s and between p.u.s and genes. The drawing is not to scale. The p.u.s are grouped into three categories on this figure: triple occurrences, double occurrences (including *his*G), single occurrence. The order of p.u.s. within each category is alphabetical. Three exceptional cases *rrn*G, *rrn*B (single) and *pro*C (double) have been represented at the bottom of the figure in the insert. ○ (resp.△) stands for a possible *rho*-independent (resp. not *rho*-independent) transcription terminator predicted from the sequence. ● (resp.▲) indicates experimental evidence for a *rho*-independent (resp. not *rho*-independent) terminator. ★ indicates experimental evidence for an RNase III processing site. ∿∿∿ ORF contained within the region represented. ····· stable RNA. **Insert. Left.** The promoter regions of the *rrn*G and *rrn*B operons. The promoters are named P_1 and P_2. The -35 and -10 regions of P_1 are indicated. The beginning of the 16S rRNA is on the right. The end of an ORF, ORFB, is indicated on the left by its translation stop codon. P.u.s. are pictured as stem and loop structures. Regions of identical sequence are drawn as continuous lines. Regions of different sequences are drawn as dotted or dashed lines. **Right**. End of gene *pro*C. P.u.s are represented as stem and loop structures. P.u.a. includes the translation stop codon for *pro*C. P.u.a. and p.u.b. overlap. The references are as follows. For triple occurrences: *aro*G (Davies and Davidson, 1982); *gdh*A (Pherson and Wootton, 1983); *gln*S (Yamao *et al.*, 1982); *gly*A (Plamann *et al.*, 1983); *lam*B (Clément and Hofnung, 1981; Gilson, 1983); *pls*B (Lightner *et al.*, 1983); *rpl*L (Barry *et al.*, 1980); *thr*C (Parsot *et al.*, 1983); *trp*R (Gunsalus and Yanofsky, 1980; Singleton *et al.*, 1980). For double occurrences: *deo*C (Valentin-Hansen *et al.*, 1984); *his*J (Higgins *et al.*, 1982b); *his*G (Higgins *et al.*, 1982a); *met*J (Zakin and Cohen, personal communication); *met*L (Duchange *et al.*, 1983); MIRNA (Reed and Altman, 1983); *ndh*-up and *ndh* (Young *et al.*, 1981); *pro*C (Deutch *et al.*, 1982); *spot*42 RNA (Rice and Dahlberg, 1982; Joyce *et al.*, 1982); $tRNA^{Gln2}$ (Nakajima *et al.*, 1981, 1982). For single occurrences: *cya* (Roy and Danchin, personal communication); *fol*A (Smith and Calvo, 1980); *his*P (Higgins *et al.*, 1982b); *lac*Y (Büchel *et al.*, 1980; Zabin and Fowler, 1970); *met*F (Saint-Girons *et al.*, 1983); *mol*A (Clément and Hofnung, 1981); *mtl*A (Lee and Saier, 1983); *orf*X-up (*lys*R) (Stragier and Patte, 1983); *phe*U ($tRNA^{Phe}$) (Schwartz *et al.*, 1983); *pho*E (Overbeeke *et al.*, 1983); *pyr*E (Poulsen *et al.*, 1983; Lundberg *et al.*, 1983); *rpo*D (Burton *et al.*, 1983); *rrn*B (Brosius *et al.*, 1981; Kingston and Chamberlin, 1981; Erdei *et al.*, 1983); *rrn*G (Shen *et al.*, 1982); *tna*A (Deeley and Yanofsky, 1981); *tsr* (Boyd *et al.*, 1983); GA (Inokuchi *et al.*, 1982).

well as extragenic and are also found in eukaryotic DNA. Interestingly, their frequency in prokaryotes is of the same order as that of the consensus stem. This confirms the idea that consensus occurrences are not random events. Finally, it should be noted that neither the consensus stem nor variations with one difference were found in the complete genome of λ or T7 phages. The few occurrences found in these phages were single, intragenic and presented at least two differences with the consensus stem (Table I).

In conclusion, the consensus stem and variations with one difference are highly characteristic of certain extragenic sequences in *E. coli* and presumably *Salmonella*. The size of the bacterial DNA data base does not allow conclusions for other bacterial species. The palindromic units and related sequences constitute a family of dispersed repetitive sequences in *E. coli*, essentially found in extragenic regions in clusters or in single occurrence.

Possible functions

Little is known on the functions of dispersed repetitive sequences in eukaryotes. The structure and location of the p.u.s in bacteria are suggestive of at least two possible roles: a role in gene expression and a role in genome evolution. We will briefly relate indications from the literature which support each of these roles. One should be careful at this stage not to exclude *a priori* other possible functions such as an involvement in DNA superstructure or nucleotide structure, or even a role in genetic transfer (Danner *et al.*, 1980). In addition there is a highly variable part in the p.u. whose function may vary accordingly.

Gene expression. It was initially suggested that p.u.s present in intergenic regions within an operon may result in a step down in expression. For example in *his*J-*his*Q, *his*G-*his*D, *lam*B-*mol*A, *mal*E-*mal*F, *rpl*L-*rpo*B, *lac*Y-*lac*A, *deo*C-*deo*A, there is evidence that expression of the distal gene is reduced

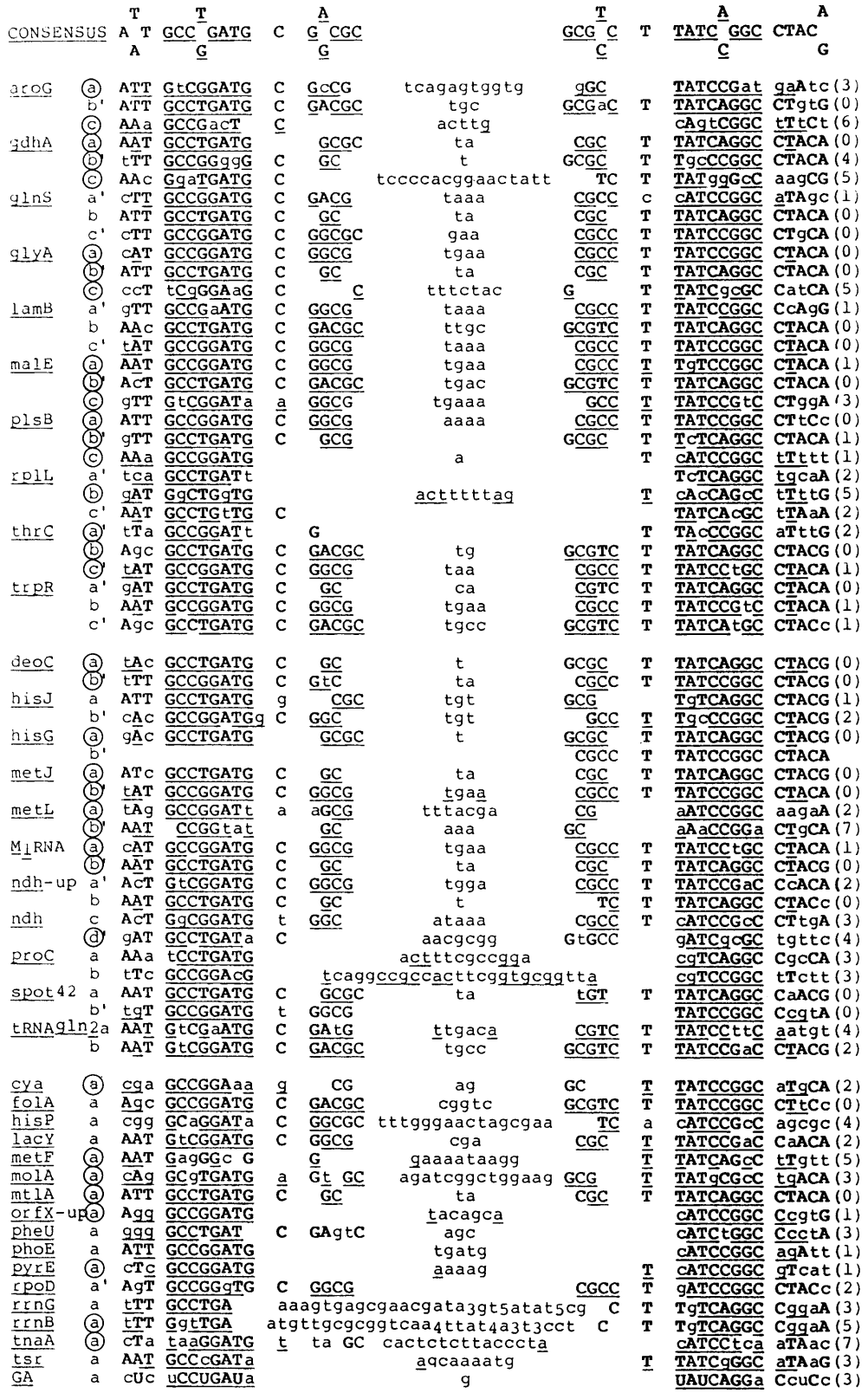

Fig. 2. Sequences of bacterial extragenic p.u.s. P.u.s are named according to the preceding gene (see Legend to Figure 1). The list gives first the clusters of three p.u.s then of two p.u.s then of single occurrences. The order is alphabetical within each category. Within each cluster the p.u.s are in the order corresponding to left to right on Figure 1. a, b, c, indicate that the sequence given is on the putative coding strand. a′, b′, c′ refer to the other strand. These letters are circled when the corresponding occurrence was found by eye search. This means either that these p.u.s were not included in the data base at the time of the search or that they did not correspond to any of the patterns looked for (see Materials and methods). P.u.s are oriented so that the 9th base of the stem is the C of the C-T mismatch or that the CTACG/A pentanucleotide is in 3′ position (see text). The nomenclature is the same as in Figure 1. The paired regions are underlined. The bases which are identical to the consensus sequence are in capital letters. The number of differences which the consensus stem is indicated between parentheses for each p.u. on the right of the figure. The free energy of formation of the stem and loop mRNA structure corresponding to the whole consensus would be -33.4 kcal $< G < -24.0$ kcal (Tinoco *et al.*, 1973) and to the 8 bp of the lower part of the consensus -22.2 kcal $< G < -18.4$ kcal.

Table I. Classification of the occurrences according to the number of differences with the consensus stem

Differences with consensus	Bacterial[a] (~143 kb)		Phages T7 and λ (~98 kb)	Eukaryotes (~1500 kb)			Bacterial
	Ex.	Int.	Int.	Ex.	Int.	?	Ex.
(0)	9	0	0	0	0	0	14
(1)	6	0	0	2	1	0	7
(2)	7	4	4	1	1	0	3
(3)	7[a]	4	1	4	5	0	3
(4)	2	1	1	3	0	3	2
(5)	–	–	–	–	–	–	5
(6)	–	–	–	–	–	–	1
(7)	–	–	–	–	–	–	2
	31	9	6	10	7	3	37
	46			20			37

[a]Including the RNA phage GA.
The figures between parentheses in the first column are the number of differences with the consensus stem. The other figures indicate the number of corresponding occurrences found. The results of the computer search (August 1983) are grouped in three categories: bacterial sequences (including the RNA phage GA), phages λ and T7, eukaryotic sequences. Bacterial extragenic sequences were examined by eye search performed until January 1984. The approximate sizes of the data base are indicated in kilobases (kb). Ex. means extragenic, i.e. in non-coding regions; Int. means intragenic, i.e. in coding regions; ? means that such information is not available; – means not determined.

compared with the proximal gene (see legend to Figure 1 for references). However the mechanism(s) involved is (are) not conclusively established. In two cases, evidence has been published. (i) In the *rplL-rpoB* region, it was shown (Barry *et al.*, 1980) that the second p.u. (*rplLb*) corresponded to a *rho*-independent terminator, while the third p.u. (*rplLc'*) included an RNase III processing site. However *rplLb* and *rplLc'* are rather atypical: the stem of *rplLb* presents five differences with the consensus stem, while *rplLc'* has no loop. It is the only cluster of three p.u.s found where a strict alternation between a GC and AT pair is not seen between successive occurrences on the fourth position of the stem. Interestingly, however, we found some homology between certain RNase III processing sites in phage T7 and the consensus stem (not shown). (ii) The case of the *deoC-deoA* intergenic region (Valentin-Hansen *et al.*, 1984) may be more relevant since the two p.u.s have a consensus stem. In this case an mRNA species with a 3' end corresponding to the end of *deoCb'* was detected. The authors suggest that it could correspond to *rho*-independent termination, but do not exclude that the p.u.s could behave as a degradation barrier for the untranslated 3' end of DNA.

In the present search we have detected several occurrences which are located at the end of operons, or between convergent operons. At the end of the *rpsU-dnaG-rpoD* operon there is a *rho*-independent terminator, just preceded by a single p.u. At the end of the *glyA* gene, there is a triple occurrence and experimental evidence suggests that termination of transcription occurs within the third p.u. (*glyAc*). There are several examples of convergent operons with p.u.s in between. In the case of *plsB* and *thrC*, a cluster of three p.u.s. is found respectively between these genes and convergent open reading frames (ORF). Whether these ORFs are expressed is not known. There is a cluster of two p.u.s located between the terminator of the *polA-spot42* operon and the 22-kd protein which is expressed convergently. The region between the *phoE* gene and the convergently expressed ORF (corresponding to a 130-kd protein) consists essentially of a single p.u. At least in this last case one may expect the p.u. to play a role in resolving the convergent mRNAs. One obvious possibility is transcription termination, but other processes involving, for example, RNA cleavage or degradation arrest are not excluded. There is one peculiar case: a single p.u. is located just after gene *pyrE* within a convergent ORF which would have its stop codon overlapping with that of the PyrE protein.

In summary, p.u.s have been identified as transcription terminators (*rplLb*, *deoCb'*, *glyAc* and possibly *folA*) or RNase III processing site (*rplLc'*). In some cases p.u.s are located just ahead of terminators (the 22-kd ORF opposite to *spot42*, *rpoD*, the ORF opposite the $tRNA^{Gln2}$). Preliminary evidence suggests that the *lamB* p.u.s do not correspond to transcription termination (Gilson, 1983). As mentioned above it is not excluded that at least some of the p.u.s identified as transcription terminators could turn out to be processing sites or degradation barriers.

Transcription termination is a complex phenomenon which often involves several DNA palindromic structures as well as several steps (Wu *et al.*, 1981; Cau *et al.*, 1982). P.u.s could play a role in some of these steps. Combination of termination structures including p.u.s could result in different mode of termination or reflect mechanisms. In particular, as suggested (Clément and Hofnung, 1981; Higgins *et al.*, 1982a), p.u.s within clusters could combine to form structures reminiscent of attenuators. In one case (*proC*) there is a system of two overlapping p.u.s one of which includes the *proC* stop codon. Finally, it is worth mentioning that in eukaryotes, transcription termination and the formation of mRNA 3' ends may be, at least in some cases, separate processes (reviewed by Proudfoot, 1984). If that is the case in prokaryotes, it is tempting to speculate that p.u.s located just ahead of terminators could play a role in the formation of the 3' end, for example by RNA cleavage or degradation arrest (cf. above).

P.u.s. and genome evolution. Mobile genetic elements are believed to play important roles in both the expression and the evolution of genomes, in prokaryotes as well as in eukaryotes. In particular, plasmids may have evolved by modular construction: IS or IS-like structures are indeed usually found at the boundaries of regions encoding certain plasmidic functions (Kopecko *et al.*, 1976). Such a modular organization may be general and may also be involved in the generation of more complex structures such as bacterial genomes.

The inverted repeat structure and high degree of conservation of the p.u.s are compatible with the idea that they could derive from a transposon. Comparison of the stem consensus sequence with the end of transposons reveals one interesting possibility for homology. IS10R and IS10L end with the sequence 5' CTGATG.....3' (Kleckner, 1981) which is included in the consensus sequence for the base of the stem of the p.u. (GCCTGATG). Moreover, among the ends of transposable elements (Kleckner, 1981) one finds a family of five nucleotide sequences related to $TGAT^{T}/G^{T}/A$ which includes five nucleotides of the consensus stem TGATG of the p.u. The p.u.s could promote rearrangements between distant regions of the genome such as the ones detected in *Salmonella* (Anderson and Roth, 1978). Single p.u.s are present at the beginning of the highly homologous *rrnB* and *rrnG* operons which encodes rRNA: interestingly these p.u.s constitute one

boundary between homologous and non-homologous regions, suggesting that they may have been a site for rearrangements.

Materials and methods

Searches for palindromic units

We performed a computer search in August 1983 and used the Sumex facilities at Stanford University and the facilities provided by Intelligenetics. We used the Quest program (Abarbanel *et al.*, 1984) and looked for the string GCC^{T}/GGATP {0,25} $QATC^{A}$/CGGC where P is purine, Q is pyrimidine. {0,25} means that the sequence between the left arm of the string GCC^{T}/GGATP and the right arm $QATC^{A}$/CGGC can include from 0 to 25 bases. The two arms correspond to the 8 bp which constitute the consensus sequence of the lower part of the stem of the p.u. which we abbreviate here as consensus stem. We also looked at all variations where one position in each arm can be any base N (N is G,T,A or C) such as: NCC^{T}/GGATP {0,25} $NATC^{A}$/CGGC; NCC^{T}/GGATP {0,25} $QNTC^{A}$/CGGC; NCC^{T}/GGATP {0,25} $QANC^{A}$/CGGC; etc... This search yielded occurrences of the p.u. which had at most four differences with the consensus stem (Table I).

We screened the NIH data base which contained at the time 1 699 705 bases corresponding to 2082 sequences belonging to 1773 loci. The *E. coli* sequences represented ~143 000 bases, i.e. ~8% of the total base and ~3% of the *E. coli* chromosome. Sequences from other bacterial species represented ~39 000 bases, i.e. ~27% of the *E. coli* sequences. In addition, we also examined with Quest the complete sequence of phage λ (48 514 bases) (Sanger *et al.*, 1983) and of phage T7 (49 936 bases) (Dunn and Studier, 1983).

After the computer search we screened, by eye, extragenic bacterial sequences which were not in the data base as well as regions around the bacterial occurrences detected with the program. This allowed the detection of a number of other occurrences some of which presented more than four differences with the consensus stem (up to seven). The summary of the computer and eye searches is given in Table I. Details are given in Results and Discussion.

Probabilities and frequencies considerations

Consensus stem. The probability of finding the consensus stem is the probability of finding GCC^{T}/GGATG or $TATC^{A}$/CGGC with their exact complement separated by at most 25 bases, namely: $(0.25)^{7} \times 0.5 \times (0.25)^{7} \times 0.5 \times 25 \times 2 = 4.7 \times 10^{-8}$. Since the size of the *E. coli* genome does not exceed 5×10^{6} bp (Bachman and Low, 1980) the probability of finding the consensus stem once in *E. coli* is ~0.24. Assuming a Poisson distribution of the p.u. among DNA sequences the probability of finding 21 occurrences of the consensus stem in *E. coli* is

$$\frac{e^{-0.24} \times (0.24)^{21}}{21!} = 1.5 \times 10^{-33}$$

Strings differing from the consensus stem. The probability of finding GCC^{T}/GGATP or $QATC^{A}$/CGGC with their exact complement separated by at most 25 bases is $2 \times (1/4)^{7} \times (1/4)^{7} \times 25 = 1.875 \times 10^{-7}$. Since the respective probabilities of finding the consensus sequence for eight nucleotides of one arm with 0, 1, 2 or 3 differences are respectively $2/4^{8}$, $46/4^{8}$, $466/4^{8}$ and $2734/4^{8}$, the probabilities of finding the consensus sequence (or its complement) followed by its complementary sequence within the next 25 bp are 2.1×10^{-6}, 4.41×10^{-5} and 5.8×10^{-4} when considering respectively at most 1, 2 or 3 differences from the 16 bases of the consensus.

Acknowledgements

We thank S.Froshauer, J.Beckwith, A.Roy, A.Danchin, M.Zakin and G.Cohen for allowing us to quote their unpublished sequence data. We thank B.Caudron and A.Buchman for help with the computer. We thank E.Dassa, P.Marliere, C.Parsot, D.Perrin and M.Zakin for useful comments on the manuscript. This work was supported by grants from the DGRST and CNRS (CP.960002, ATP.956144), the NATO (grant 1297), the Fondation pour la Recherche Médicale, the Ligue Nationale Française contre le Cancer and the Association pour le Développment de la Recherche sur le Cancer. D.B. was supported by an NIH senior Fogarty international fellowship during the initial phase of this work.

References

Arbarbanel,R.M., Wieneke,P.R., Mansfield,E., Jaffe,D.A. and Brutlag, D.L. (1984) *Nucleic Acids Res.*, **12**, 263-280.

Anderson,R.P. and Roth,J.R. (1978) *Cold Spring Harbor Symp. Quant. Biol.*, **43**, 0-0.

Bachman,B.J. and Low,B.K. (1980) *Microbiol. Rev.*, **44**, 1-56.

Barry,G., Squires,C. and Squires,C.L. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 3331-3335.

Boyd,A., Kendall,K. and Simon,M.I. (1983) *Nature*, **301**, 623-627.

Brosius,J., Dull,T.J., Sleeter,D.D. and Noller,H.F. (1981) *J. Mol. Biol.*, **148**, 107-127.

Büchel,D.E., Gronenborn,B. and Müller-Hill,B. (1980) *Nature*, **283**, 541-545.

Burton,Z.F., Gross,C.A., Watanabe,K.K. and Burgess,R.R. (1983) *Cell*, **32**, 335-349.

Cau,L., Roberts,J. and Wu,R. (1982) *Nature*, **283**, 541-545.

Clément,J.M. and Hofnung,M. (1981) *Cell*, **27**, 507-514.

Danner,D., Deich,R., Sisco,K. and Smith,H. (1980) *Gene*, **11**, 311-318.

Davies,W.D. and Davidson,B.E. (1982) *Nucleic Acids Res.*, **10**, 4045-4058.

Deeley,M.C. and Yanofsky,C. (1981) *J. Bacteriol.*, **147**, 787-796.

Deutch,A.H., Smith,C.J., Rushlow,K.E. and Kretschmer,P.J. (1982) *Nucleic Acids Res.*, **10**, 7701-7710.

Duchange,N., Zakin,M.M., Ferrara,P., Saint-Girons,I., Park,I., Tran,S.V., Py,M.C. and Cohen,G.N. (1983) *J. Biol. Chem.*, **258**, 14868-14873.

Dunn,J.J. and Studier,F.W. (1983) *J. Mol. Biol.*, **166**, 477-535.

Erdei,S., Boros,I., Szabo,G. and Venetianer,P. (1983) *Mol. Gen. Genet.*, **191**, 162-164.

Gilson,E. (1983) Thèse de 3ème cycle, University of Paris VII.

Gunsalus,R.P. and Yanofsky,C. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 7117-7121.

Higgins,C.F., Ferro-Luzzi Ames,G., Barnes,W.M., Clément,J.M. and Hofnung,M. (1982a) *Nature*, **298**, 760-761.

Higgins,C.F., Haag,P.D., Nikaido,K., Ardeshis,F., Garcia,G. and Ferro-Luzzi Ames,G. (1982b) *Nature*, **298**, 723-727.

Inokuchi,Y., Hirashima,A. and Watanabe,I. (1982) *J. Mol. Biol.*, **158**, 711-730.

Joyce,C.M., Kelley,W.S. and Grindley,N.D.F. (1982) *J. Biol. Chem.*, **257**, 1958-1964.

Kingston,R. and Chamberlin,M.J. (1981) *Cell*, **27**, 523-531.

Kleckner,N. (1981) *Annu. Rev. Genet.*, **15**, 341-404.

Kopecko,D.J., Brevet,J. and Cohen,S.N. (1976) *J. Mol. Biol.*, **108**, 333-360.

Lee,C.A. and Saier,M.H. (1983) *J. Biol. Chem.*, **258**, 10761-10767.

Lightner,V.A., Bell,R.M. and Modrich,P. (1983) *J. Biol. Chem.*, **258**, 10856-10861.

Lundberg,L.G., Thoresson,H.O., Karlström,O.H. and Nyman,P.O. (1983) *EMBO J.*, **2**, 967-971.

Nakajima,N., Ozeki,H. and Shimura,Y. (1981) *Cell*, **23**, 239-249.

Nakajima,N., Ozeki,H. and Shimura,Y. (1982) *J. Biol. Chem.*, **257**, 11113-11120.

Overbeeke,N., Bergmans,H., Mansfeld,F.V. and Lugtenberg,B. (1983) *J. Mol. Biol.*, **163**, 513-532.

Parsot,C., Cossart,P., Saint-Girons,I. and Cohen,G.N. (1983) *Nucleic Acids Res.*, **11**, 7331-7345.

Pherson,M.J.Mc and Wootton,J.C. (1983) *Nucleic Acids Res.*, **11**, 5257-5266.

Plamann,M.D., Stauffer,L.T., Urbanowski,M.L. and Stauffer,G.V. (1983) *Nucleic Acids Res.*, **11**, 2065-2075.

Poulsen,P., Jensen,K.F., Valentin-Hansen,P., Carlsson,P. and Lundberg, L.G. (1983) *Eur. J. Biochem.*, **135**, 223-229.

Proudfoot,N. (1984) *Nature*, **307**, 412-413.

Reed,E.E. and Altman,S. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 5359-5363.

Rice,P.W. and Dahlberg,J.E. (1982) *J. Bacteriol.*, **152**, 1196-1210.

Saint-Girons,I., Duchange,N., Zakin,M.M., Park,I., Margarita,D., Ferrara, P. and Cohen,G.N. (1983) *Nucleic Acids Res.*, **11**, 6723-6732.

Sanger,F., Coulson,A.R., Hong,G.F., Hill,D.F. and Petersen,G.B. (1983) *J. Mol. Biol.*, **162**, 729-773.

Schwartz,I., Klostsky,R.A., Elseviers,D., Gallagher,P.J., Krauskopf,M., Siddiqui,M.A.Q., Wong,J.F.H. and Roe,B.A. (1983) *Nucleic Acids Res.*, **11**, 4379-4389.

Shen,W.F., Squires,C. and Squires,C.L. (1982) *Nucleic Acids Res.*, **10**, 3303-3313.

Singleton,C.K., Roeder,W.D., Bogosian,G., Sommervill,R.L. and Weith, H.L. (1980) *Nucleic Acids Res.*, **8**, 1551-1560.

Smith,D.R. and Calvo,J.M. (1980) *Nucleic Acids Res.*, **8**, 2255-2274.

Stragier,P. and Patte,J.C. (1983) *J. Mol. Biol.*, **168**, 333-350.

Tinoco,I., Borer,P.N., Dengler,B., Levine,M.D., Uhlenbeck,O.C., Crothers, D.M. and Gralla,J. (1973) *Nature New Biol.*, **246**, 40-41.

Valentin-Hansen,P., Hammer-Jespersen,K., Boetius,F. and Svendsen,I. (1984) *EMBO J.*, **3**, 179-183.

Wu,A., Christie,G. and Platt,T. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 2913-2917.

Yamao,F., Inokuchi,H., Cheung,A., Ozeki,H. and Söll,D. (1982) *J. Biol. Chem.*, **257**, 11639-11643.

Young,I.G., Rogers,B.L., Campbell,H.D., Jaworowski,A. and Shaw,D.C. (1981) *Eur. J. Biochem.*, **116**, 165-170.

Zabin,I. and Fowler,A. (1970) in *The Lactose Operon*, published by Cold Spring Harbor Laboratory Press, NY, pp. 27-46.

Received on 4 November 1983; revised on 20 March 1984