Proc. Nat. Acad. Sci. USA
Vol. 70, No. 9, pp. 2667-2670, September 1973

Alignment of Two DNA Helices: A Model for Recognition of DNA Base Sequences by the Termini-Generating Enzymes of Phage λ, 186, and P2

(cohesive ends/sequence recognition/DNA alignment)

JAMES C. WANG AND DONALD P. BREZINSKI

Department of Chemistry and Group in Biophysics, University of California, Berkeley, Calif. 94720

Communicated by A. D. Kaiser, June 4, 1973

ABSTRACT **Based on the 3′- and 5′-terminal sequences of DNA of phage λ, P2, and 186, a model is proposed for recognition of DNA sequences by enzymes responsible for generation of cohesive ends. Two copies of the cohered ends, either on separate molecules or on a concatemer, are aligned with their helical axes parallel but running in opposite directions. The nicking system is dimeric, with each of the two monomers carrying identical sequence-recognition sites. Two pairs of nicks are introduced into the two aligned DNA molecules by the nicking system. The applicability of this model to other biological processes, such as integration of a viral genome into a host genome and the cutting of concatemeric T7 DNA, is discussed.**

For several phages, the DNA extracted from mature phage has short, single-stranded ends with complementary base sequences (1). For coliphages, it appears that there are only two classes of cohesive ends, with those of the lambdoid phages (λ, φ80, 21, 82, 424, and 434) in one class and those of the 186 family (including 186, P2, P4, and 299) in another (1, 2). Within each class base sequences of the ends are identical or very similar. The base sequences of λ-type cohesive ends and 186-type cohesive ends are different, and mutual joining does not occur. Shortly after infection, the cohesive ends join and the two single-chain interruptions in the resulting molecule are subsequently sealed by ligase (1). Therefore, in order to regenerate the single-stranded ends before maturation of the phage, whether from a concatemer or from a circular DNA, two staggered single-chain scissions must be introduced at unique points of the DNA molecule. The enzyme or enzymes involved in this process has been named the termini-generating enzyme or *Ter* for short (3, 4). For λ, it appears that the phage gene-A product provides the *Ter* function, although participation of host functions and λ genes between R and A has not been ruled out (5).

We studied the problem of DNA sequence recognition involved in the *Ter* system, aside from interests in this system itself, because the DNA sequences flanking the sites of endonucleolytic nicking are present at the 3′ and 5′ termini of mature phage DNA; therefore, the base sequences can be determined without the formidable difficulties of determining the sequence of a segment in the middle of a DNA molecule. Elegant work on determination of the sequence of the 5′ single-stranded ends has been done by Wu *et al.* (6, 7) and by Murray and Murray (2). Information on the 3′-termini sequences comes primarily from the work of Weigel *et al.* (8) and our own effort (ref. 9; D. P. Brezinski and J. C. Wang, unpublished).

In this communication, we postulate a model for recognition of DNA base sequences by *Ter*, based on sequence information. We believe that sequence recognition in the *Ter* system involves alignment of two copies of the cohered ends with their helical axes parallel but running in opposite directions. We further postulate that the alignment is achieved by a sequence-specific protein subunit that can dimerize, and that endonucleolytic cutting is done by a catalytic subunit that positions itself by interacting with the sequence-recognition subunit. Therefore, the model suggests that the *Ter* system may have features shared by many other genetic processes, such as integration of a viral genome into a host genome and the cutting of concatemeric T7 DNA.

RESULTS AND DISCUSSION

Sequence information

The base sequences known in the region of the cohered ends of DNA of λ, 186, and P2 are shown in Fig. 1. The phosphodiester bonds at which single-chain scissions are to be introduced are depicted as $(3'x_15')$ and $(5'x_23')$. Endonucleolytic attack at these sites gives the 3′-hydroxyl and 5′-phosphoryl termini of mature DNA.

Several observations can be made with regard to recognition of nicking sites by the *Ter* enzyme. First, for a given DNA, the sequences flanking the nicking sites x_1 and x_2 are different. For λ, the sequences adjacent to the nicking sites are ...CG-$(3'x_15')$AG... and ...GG$(5'x_23')$GC..., respectively. The corresponding sequences are ...CA$(3'x_15')$AT... and ...GG$(5'x_23')$AC for 186, and ...CA$(3'x_15')$GT... and ...GG$(5'x_23')$GG... for P2.

While for λ and 186 DNA the 3′ sides of the sequences appear to be related by a symmetry element (...CG$3'x_1$ and x_23'GC... for λ and ...CA$3'x_1$ and x_23'AC... for 186), this is not so for P2 (...CA$3'x_1$ and x_23'GG...). Second, the sequences of the cohesive ends of mature 186 and P2 DNA are rather similar. Only two out of 19 base pairs are different: an AT → GC substitution to the immediate right of x_1 and an AT → TA inversion five base pairs to the left of x_2. However, when the sequences adjacent to the sites of nicking are examined, the difference is fairly large. Two bases differ at x_2 [...GG$(5'x_23')$GG... for P2 and ...GG$(5'x_23')$AC... for 186], and one base differs at x_1 [...CA$(3'x_15')$GT... for P2] and [...CA$(3'x_15')$AT... for 186].

We proposed previously that the *Ter* functions of 186 and P2 are likely to be similar to the extent that the *Ter* enzyme of one can cut the DNA of the other (10). This sug-

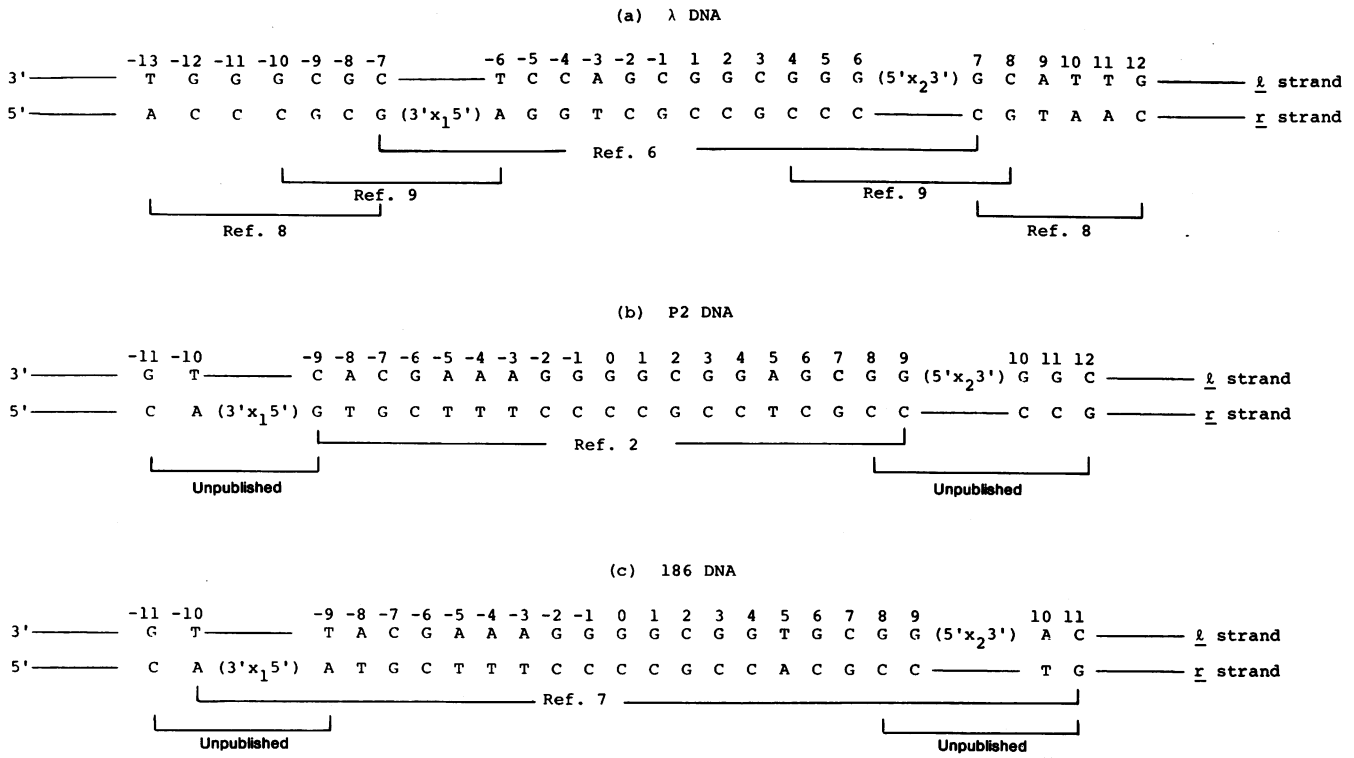

FIG. 1. Base sequences in the cohered-ends region of DNA of λ, P2, and 186. The two phosphodiester bonds at which endonucleolytic nicking is to occur are indicated as $(3'x_1 5')$ and $(5'x_2 3')$. The base pairs are numbered from the position midway between the two nicking points. *Unpublished* = D. P. Brezinski and J. C. Wang.

gestion was based on the fact that 186 can serve as helper for the satellite phage P4, which has cohesive ends identical to those of P2. Since P4 has few late genes, it probably uses the *Ter* function of the helper 186 to generate its cohesive ends.

The nonidentity of the sequences at the cutting sites, for either the two sites of one molecule or when the corresponding sites of P2 and 186 DNA are compared, implies that the cutting system does not recognize sequences immediately adjacent to the cutting sites. Therefore, it appears that the *Ter* enzyme binds to a recognition site or sites, and the positions of nicking are determined by the relative positions of the catalytic site and the DNA sequence-recognition site of the enzyme. For λ DNA, Weigel *et al.* have pointed out that there is a 2-fold axis of symmetry (8), located midway between the two nicks, with regard to purines and pyrimidines. This symmetry is not perfect; the second base pair on each side of the central plane, for example, does not follow this rule, and the term "hyphenated symmetry" was used.

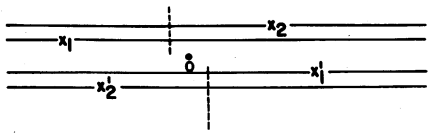

FIG. 2. A schematic drawing showing two copies of the cohered ends aligned with their helical axes parallel but running in opposite directions. The two sites of endonucleolytic attack are shown as x_1 and x_2 for the upper molecules and x'_1 and x'_2 for the lower molecule. *Dotted lines* are the planes midway between x_1 and x_2 and between x'_1 and x'_2. There is not sufficient information to specify the spacing between these two planes. Point 0 is the center of symmetry in this diagram.

Since the two nicking sites on the DNA are related by this "hyphenated symmetry," it might be sufficient for the *Ter* enzyme of λ to recognize only half of the base sequence in the cohered-ends region. For 186 DNA and P2 DNA, however, such "hyphenated rotational symmetry" is not present, as noted by Weigel *et al.* (8).* We are therefore left with two alternatives. Either the sequences for recognition by the *Ter* enzymes of P2 and 186 are outside the sequences shown in Fig. 1, or the sequences for recognition are in the cohered-ends region and there are two catalytic sites on the enzyme properly spaced. The latter alternative requires the two catalytic sites on the enzyme to be about 70 Å apart. This distance is fairly large, although it is not outside the range for a large protein molecule.

As an alternative to models involving recognition of one copy of the end join, we have considered a model that involves two copies of the cohered ends.

The model

Essential features of the model are as follows: (*i*) Sequence recognition in the *Ter* system involves alignment of *two*

* For the linear arrays of sequences shown in Fig. 1, we note that for P2 and 186 there is a "hyphenated mirror symmetry" with the mirror plane passing through the GC pair at position 0. In the region between the two nicking sites, exceptions to this symmetry (hyphens) occur at positions ±2 and ±9 for P2 and ±2, ±5, and ±9 for 186. The change from the "hyphenated mirror symmetry" in P2 to the "hyphenated rotational symmetry" in λ can be viewed as a result of a switching of the sequence 5'CCTCGC in the r-strand of P2 to the l-strand in λ.

copies of cohered ends, either on separate molecules or on a concatemer, with their helical axes parallel but running in opposite directions. The alignment brings a left-end nicking site on one copy of the end join, x_1, and a right-end nicking site on the other copy of the end join, x_2', into close proximity. Furthermore, because of the symmetry introduced by this alignment, the stereochemical environment at one pair of nicking sites, x_1 and x_2', is the same as the stereochemical environment at the other pair of sites, x_2 and x_1' (Fig. 2).

(*ii*) Proper alignment is achieved by a protein ϕ, which has a DNA sequence-recognition site and a dimerization site. This hypothesized protein may be either a viral or host factor.

(*iii*) The catalytic unit of the *Ter* system positions itself by specific interactions with ϕ. Each pair of nicks, x_1 and x_2' or x_2 and x_1', are introduced more or less in concert. For each pair of aligned helices, the two pairs of nicks could be introduced either by one catalytic unit sequentially, or perhaps more likely, by two identical catalytic units, one on each side of the dimerized ϕ molecules.

Further discussion

The sequences listed in Fig. 1 suggest that the *Ter* system of λ and that of P2 and 186 might be related. A comparison between the sequences of λ and P2 shows that the sequences CG(3′$x_2$5′)GGCG and 5′GGCGG in the l-strand of λ (positions 8–2 and 5–1, respectively) are also present in the l-strand of P2 (positions 12–6 and 4–0, respectively). Furthermore, continuing on the l-strand of λ, the sequence 5′CCTCGC (positions −4 to −9) is present in the *r-strand* of P2 (positions 3–8). These similarities are unlikely to be due to random chance, and therefore suggest that the *Ter* systems of all these phages are related: either they are evolved from a common ancestral protein or they involve a common host function. These possibilities can be tested experimentally. If a common host function is involved, then there might be a class of host mutants of the *gro* type discussed by Georgopoulos and Herskowitz (11) and by Georgopoulos (12). Such "*gro Ter*" mutants would not support the growth of normal phage with either the λ-type or the 186-type cohesive ends. If the viral functions are similar, since gene-A product of λ is or is not part of the *Ter* system, heteroduplex mapping (13) must reveal significant sequence homology between the A gene of λ and a corresponding gene in P2 and 186 DNA.

Our model also makes a definitive prediction on the dependence of *Ter* function on the concentration of cohered ends. The seemingly puzzling observation that *in vivo Ter* appears to require more than one copy of the viral genome (3) is a natural consequence of the model.

In our model, *Ter* requires two copies of cohered ends. The two copies can be on one DNA molecule, either a circular dimer or a concatemer, or on two separate molecules such as circular monomers. It does not necessarily follow, though, that in each case the resulting monomeric DNA can be packaged into a phage particle (Fig. 3). Fig. 3*a* depicts a circular dimer with the two end joins XY and Y′X′ aligned. It is plausible that the event of cutting by *Ter* and packaging of the DNA are temporally related. Each pair of molecular ends generated by *Ter*, X and Y′ and Y and X′, might remain in close proximity during packaging as well, by serving for example, as the nucleation site for packaging. This would result in the successful packaging of at least one, and probably two, monomeric mature DNA molecules. It is easy to see that for two separate monomeric DNA molecules (Fig. 3*b*), the mechanism described would not be able to package the DNA molecules even though the cutting can be done by *Ter* (14).†

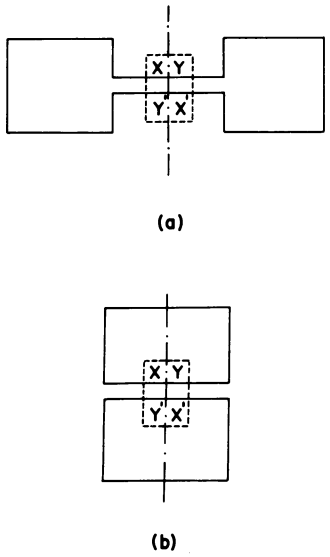

FIG. 3. A schematic drawing demonstrating a difference between two copies of the end joins, XY and X′Y′, on the same molecule (*a*), and on different molecules (*b*). (*a*) A dimeric ring is shown, packaging of the monomers resulting in X and Y′ being in one phage particle and X′ and Y in the other. (*b*) Two separate monomeric rings are shown. It is not possible to have X and Y′ or X′ and Y go into one phage particle.

Since the protein(s) involved in the *Ter* system is not isolated, it is speculative at this time as to which sequences are involved in the recognition. We suspect that the pentameric sequence $\binom{5'GGCGG}{3'CCGCC}$, which is present in all three end joins examined, might be one of the elements involved in recognition.‡

Alignment of DNA molecules occurs in many important biological processes, such as prophage integration, genetic recombination, and chromosomal pairing. The model described for the *Ter* system has several features that might be applicable to other pairing and cutting problems. Our model does not involve Watson–Crick type pairing between the two molecules. This may have bearing on the prophage

† This might be the reason that unintegrated λ chromosome in monomeric form cannot be packaged by phage ϕ80 or 21.

‡ For the cohered-ends region of P2 and 186, we note that there is a clustering of purines in one strand and pyrimidines in the other (18 purines out of 24 shown are in one strand). Therefore, when two copies of the sequences are aligned in parallel but opposite direction, the probability of a purine–purine and pyrimidine–pyrimidine matching between the two l-strands or the two r-strands of the two helices is high. If the two nicking sites x_1 and x_2' are aligned, "hyphenated mirror symmetry" results in a purine–purine and pyrimidine–pyrimidine matching between the two l-strands or r-strands (or a purine–pyrimidine matching between an l-strand and an r-strand), with exceptions only at the positions of the "hyphens." It is conceivable that such matching may play a role in alignment of helices. The significance of the "hyphenated symmetry" in λ for recognition by *Ter*, if any, is not apparent in our model.

integration problem. Here the viral and bacterial attachment sites must have some "homology," yet no Watson–Crick type pairing has been detected (15). Whether our model is directly applicable in prophage integration will become clear when the base sequences of the attachment sites are known.

The analogy between the cutting of the cohesive ends discussed in this communication and the cutting of concatemeric T7 DNA is apparent, and it is possible that DNA alignment is also involved in the latter. We believe that the essential features of our model, for the *Ter* systems in particular and for other biological processes of similar nature, will stand further experimental tests.

We thank Drs. Englund and K. Murray for communicating to us their base sequence results prior to publication. This work has been supported by grants from the U.S. Public Health Service (GM14261) and the National Science Foundation (GB29331). D.P.B. has been supported by an NSF Predoctoral Fellowship and a U.S. Public Health Service Training Grant 5T01-GM-00829-11.

1. For a review, see Kaiser, A. D. & Wu, R. (1968) *Cold Spring Harbor Symp. Quan. Biol.* **33,** 729–734.
2. Murray, K. & Murray, N. E. (1973) *Nature,* in press.
3. For a review, see Yarmolinsky, M. B. (1971) in *The Bacteriophage Lambda,* ed. Hershey, A. D. (Cold Spring Harbor Laboratory, New York), pp. 97–111.
4. Mousset, S. & Thomas, R. (1969) *Nature* **221,** 242–244.
5. Wang, J. C. & Kaiser, A. D. (1973) *Nature New Biol.* **241,** 16–17.
6. Wu, R. & Taylor, E. (1971) *J. Mol. Biol.* **57,** 491–511.
7. Padmanabhan, R. & Wu, R. (1972) *J. Mol. Biol.* **65,** 447–467; Wu, R., Donelson, J., Padmanabhan, R. & Hamilton, R. (1972) *Bull. L'Inst. Pasteur* **70,** 203–233.
8. Weigel, P. H., Englund, P. T., Murray, K. & Old, R. W. (1973) *Proc. Nat. Acad. Sci. USA* **70,** 1151–1155.
9. Brezinski, D. P. & Wang, J. C. (1973) *Biochem. Biophys. Res. Commun.* **50,** 298–404.
10. Wang, J. C., Martin, K. & Calendar, R. (1973) *Biochemistry,* in press.
11. Georgopoulos, C. P., & Herskowitz, I. (1971) in *The Bacteriophage Lambda,* ed. Hershey, A. D. (Cold Spring Harbor Laboratory, New York), pp. 553–564.
12. Georgopoulos, C. P., (1971) in *The Bacteriophage Lambda,* ed. Hershey A. D. (Cold Spring Harbor Laboratory, New York), pp. 639–645.
13. Davis, R. W., Simon, M. & Davidson, N. (1971) in *Methods in Enzymology,* eds. Grossman, L. & Moldave, K. (Academic Press, New York), Vol. XXI, pp. 413–428.
14. Dove, W. F. (1968) *Annu. Rev. Genet.* **2,** 305–340; Szpirer, J. & Brachet, P. (1970) *Mol. Gen. Genet.* **108,** 78–92.
15. Davis, R. W. & Parkinson, J. S. (1971) *J. Mol. Biol.* **56,** 403–423.