SHORT PAPER

Tn3 resolvase-like sequence in *P* transposable element of *Drosophila melanogaster*

BY Yōko SATTA*, Takashi GOJOBORI**, Takeo MARUYAMA**
and Sadao I. CHIGUSA*

**Department of Biology, Ochanomizu University, Tokyo 112
and*

***National Institute of Genetics, Mishima 411*

(Received May 15, 1985)

ABSTRACT

Using each one of the four open reading frames in the *P* transposable element of *Drosophila melanogaster*, computer search of homology was carried out against the protein sequence data base of NBRF version 4. A deduced polypeptide of about 210 residues in ORF1 of the *P* element appears to have a significant homology with the resolvases of bacterial transposable elements Tn3 and $\gamma\delta$. The polypeptide seems to possess a region consisting of 22 amino acid residues having homology with DNA-binding fold of the resolvases of the bacterial transposons Tn3 and $\gamma\delta$, which are known to have a repressor function of the transposable elements. There are a nine base pair region in ORF3 and also in the 31 base pair inverted terminal repeat of the *P* element having homology with a site in the promoter region of the Tn3 and also of $\gamma\delta$ at which the resolvases bind and suppress transcription of transposases. Based on these findings, it was suggested that the ORF1 might code a resolvase protein which mediates site-specific recombination and also suppresses the transposition of the *P* by binding at a specific site in the ORF3 or in the terminal repeat. It was also suggested that the whole *P* might code a transposase.

1. INTRODUCTION

The *P* transposable elements which cause hybrid dysgenesis in certain interstrain crosses of *Drosophila melanogaster* have been sequenced by O'Hare and Rubin (1983). Among *P* elements of various sizes, the one which appears to be functionally complete for causing the hybrid dysgenesis is about 2.9 kb long and has four apparent open reading frames, though proteins are not identified for them.

We have made an attempt to identify homology between the hypothetical polypeptide sequence deduced from each of the four open reading frames and proteins whose primary structures are known and available in a data base. We used the data base of NBRF (National Biomedical Research Foundation, Georgetown University Medical Center) version 4 in which about 3000 protein

sequences are available for this kind comparison, though not all of them are distinct proteins. We report here the results of the analysis and discuss them briefly.

2. ANALYSIS AND RESULTS

Among the proteins at present compiled in the data base, none shows strong homology to any of the possible open reading frames of the *P* element. Although homology is weak, the one of the open reading frames is partially homologous to the resolvase proteins for *Tn3* and $\gamma\delta$ of bacterial transposable elements, which are known to mediate site-specific recombination and to suppress transcription of the transposase gene (Heffron *et al.* 1979; Reed *et al.* 1982; Neuman and Grindly 1984). Following O'Hare and Rubin (1983), we will refer to the open reading frame and the polypeptide deduced from it as ORF1.

The part of ORF1 having homology with the resolvases of the bacterial transposable elements has about 210 amino acid residues, while the corresponding regions of the resolvases are slightly shorter. When an alignment is made between the ORF1 and the resolvases, four blocks of insertion gaps appear in the ORF1 which do not exist in the resolvases while one gap is required for the resolvases to match with the ORF1. The relationship is illustrated in Fig. 1 where open boxes represent insertion gaps, and corresponding regions of the ORF1 and the resolvases are indicated by the same pattern of shading.

Assuming tentatively that the regional correspondences presented in Fig. 1 are appropriate, we made a detailed alignment between three polypeptides, ORF1 and the two resolvases. We used two different criteria for identifying the homology at each corresponding residue. One is of course identity of corresponding amino acids. The other is to detect similarity between compared polypeptides. Certain amino acid groups are known to be chemically and functionally similar, and therefore substituting one for another of the same group has little biological effect. It is further known that such substitutions between similar amino acids have occurred more frequently than among different groups in evolution. In the present paper, we adopted the method of classifying the amino acids into the four similarity groups; acidic, basic, un-

Fig. 1. Diagram indicating relationship between the ORF1 of the *P* and the bacterial resolvase.

```

Resol. (  $\gamma\delta$  )  MRLFGYARVSTSQSSL          DIQVRALKDAGVKANRIFTDKAS
Resol. (Tn 3)  MRIFGYARVSTSQSSL          DIQIRALKDAGVKANRIFTDKAS
ORF1 ( P )    MRRRLLEQQLLRESQQLLEESLRKIFTDTQIRILKNGGQRATFFNSDDISTAICLHTAGPRAYNHLYKKGFP

Resol. (  $\gamma\delta$  )  GSSSDRKGLDLLRMKVEEGDVILVKKLDRLGRDTADMQLIKEFDAQGVSIIRFIDDGISTD  GEMGK
Resol. (Tn 3)  GSSTDRLEGLDLLRMKVEEGDVILVKKLDRLGRDTADMQLIKEFDAQGVAVRFIDDGISTD  GDMGQ
ORF1 ( P )    LPSRTILYRWLSDVDIKRGCLDVVIDLMDSDGVDDADKLCVLAFDDEMKVAAAFEYDS  SADIVYEPS

Resol. (  $\gamma\delta$  )  MVTILSAVAQAERORILERTNEGRQEAMAKGVVFGRRKRKIDR          DAVLNMWQOG
Resol. (Tn 3)  MVTILSAVAQAERRRILERTNEGRQEAKLKGIFGRRRTVDR          NVVLTTHQKG
ORF1 ( P )    DYVQLAIVRGLKKSWKQPFFFDFNTRMDPDTLNNILRLHRLKGYLVVA

Resol. (  $\gamma\delta$  )  LGASHTSKTMNIAIRSTVYKVINEN  SN
Resol. (Tn 3)  TGATEIAHQLSIAIRSTVYKILED  ERAS
ORF1 ( P )    IVSDLGTGNQKLWTELGISECKFRILOKSDNP

```

Fig. 2. Sequences similarity between the resolvases and the ORF1. Residue identity is indicated by solid lines and residues having similarity by broken lines. Single letter abbreviations for the amino acid residues are as follows: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

charged polar and nonpolar (cf. Alberts *et al.* 1983). Varying the criteria for the similarity, we can classify amino acids into groups in different ways (Schwartz and Dayhoff 1978; Miyata *et al.* 1979; Alberts *et al.* 1983). However, when these methods were applied, conclusions remained much the same and hence the results were not presented here.

Homology based on the strict identity and similarity as defined above are shown in Fig. 2. In the figure, amino acid residues being identical to the corresponding residues of the resolvases are enclosed by solid lines and residues being similar to the resolvase of Tn β by the above criterion are enclosed by broken lines.

3. DISCUSSION

These two resolvases of the bacterial transposable elements have strong homology and are functionally interchangeable (Chou *et al.* 1979; Reed 1981; Reed *et al.* 1982). The present analysis of homology between the open reading frames of the *P* and the bacterial resolvases suggests that the ORF1 might code a resolvase-like protein with dual functions. Presently available information on the mechanism of the site-specific recombination does not allow further inquiry into the relationship between the bacterial and *Drosophila* systems. However, as for the suppressor function of the resolvases, we can speculate further. It is known that each DNA-binding protein has a specific string of about 20 residues being responsible for the binding, and that these strings usually form alpha helices and share a common feature of having some characteristic amino acids (Pabo and Sauer 1984; Abdel-Meguid 1984). Furthermore these DNA-binding proteins bind at a special DNA site having characteristic base pairs. A typical site has nine base pairs of sequence TGTCYNNTA from 5' to 3' termini. Therefore we first looked for such a string in the ORF1 and found a possible candidate which appears to have the common feature of the DNA-binding proteins. The region showing apparent homology to the DNA-binding folds consists of 22 amino acid residues of TGNQKLWTELGISECKFRILQK in which several residues are of particular importance and shared by most of 16 DNA-binding proteins examined, including the two bacterial resolvases being discussed here (Laughon and Scott 1984). As for the presumed DNA-binding site, we found a region of nine bases of sequence TGTCTTTTA in the ORF3 having corresponding bases of known DNA-binding targets. Therefore the analysis enables us to suggest that one of the open reading frames, ORF1, of the *P* element may code a dual enzyme whose partial function is to suppress the action of the *P* under a certain circumstance. It further suggests that the suppression may occur by acting at a specific site in the ORF3. If the region in the ORF3 and the protein encoded by the ORF1 are in fact responsible for the suppression of the trans-

position of the *P* element, the analogy between the *P* system and the bacterial system holds for both the resolvases and DNA-binding mechanism. This may allow us to speculate further that, by the analogy of the two systems of transposable elements, the ORF3 or the whole *P* might indeed code a transposase of the *P*. The transposase of Tn3 is 1015 residue long (Heffron *et al.* 1979), while the entire *P* can code approximately 900 amino acids. Furthermore, poly-A RNA corresponding to the entire *P* is obtained and mutations induced in different ORFs do not complement each other (Karess and Rubin 1984). These evidence suggests that the whole *P* may code a transposase. Therefore we compared the theoretical polypeptides of the *P* with the Tn3 transposase. Weak homology appears to exist between them. The ORF0 of the *P* with 99 residues appears to be homologous to a region starting around the 60-th residue of the Tn3 transposase, and furthermore the ORF1, 2, and 3 appear to be also weakly homologous to regions down stream of the Tn3 in this order. We also found a rather striking feature of a 61 base pair intron between the ORF2 and ORF3. When this intron region is translated and compared with the Tn3, there is weak homology at a region around the 750-th residue of the Tn3.

We also found a region of DNA in the terminal repeat of the *P*, showing a slightly weaker homology to the known DNA binding targets. This may lead to a possible difference between the two systems. In the case of the bacterial system, the resolvases bind at a nucleotide sequence in the promoter region and suppresses the transcription. On the other hand, it is known that the transcription of the *P* element takes place in the system where intact *P* elements are present, but the dysgenesis is suppressed by a certain cytotype (Karess and Rubin 1984). Thus the apparent analogy between the bacterial and the *Drosophila* *P* systems does not go this far. Hence it is possible that if the site in the terminal repeat is responsible for the suppression, the presumed protein encoded by the ORF1 binds at both terminal repeats of the *P* and blocks the transposases produced by the *P*. This will hinder the transposition of the *P* element by the transposases.

We are grateful to Dr. Y. Sano and Dr. J. F. Crow for their comments. This study was partially supported by Grants-in-aid for Scientific Research No. 57120001, 57540359, and 58380020 from the Ministry of Education, Science and Culture.

Note added in proof; Dr. K. Saigo communicated us that he had aligned ORF2-3 with Tn3 transposase.

REFERENCES

- ABDEL-MEGUID, S. S., GRINDLEY, N. D. F., TEMPLETON, N. S. and STEITZ, T. S. (1984) Cleavage of the site-specific recombination protein $\gamma\delta$ resolvase: The smaller of two fragments binds DNA specifically. *Proc. Natl. Acad. Sci. USA* **81**, 2001-2005.
- ALBERTS, B., BRAY, D., LEWIS, J., RAFF, M., ROBERTS, K. and WATSON, J. D. (1983) *Molecular Biology of the Cell. Chapt. 2*. Garland Publishing, New York and London

- CHOU, J., CASADABAN, M. J., LEMAUX, P. G. and COHEN, S. N. (1979) Identification and characterization of a self-regulated repressor of translocation of the Tn3 element. *Proc. Natl. Acad. Sci. USA* **76**, 4010-4024.
- GRINDLEY, N. D. F., LAUTH, M. R., WELLS, R. G., WITYK, R. J., SALVO, J. J. and REED, R. R. (1982) Transposon-mediated site-specific recombination: Identification of three binding sites for resolvase at the *res* sites of and Tn3. *Cell* **30**, 19-27.
- HEFFRON, F., MCCARTHY, B. J., OHTSUBO, H. and OHTSUBO, E. (1979) DNA sequence analysis of the transposon Tn3: Three genes and three sites involved in transposition of Tn3. *Cell* **18**, 1153-1163.
- KARESS, R. E. and RUBIN, G. M. (1984) Analysis of P transposable element functions in *Drosophila*. *Cell* **38**, 135-146.
- LAUGHON, A. and SCOTT, M. P. (1984) Sequence of a *Drosophila* segmentation gene: Protein structure homology with DNA-binding proteins. *Nature* **310**, 25-31.
- MIYATA, T., MIYAZAWA, S. and YASUNAGA, T. (1979) Two types of amino acid substitution in protein evolution. *J. Molecular Evolution* **12**, 219-236.
- NEUMAN, B. J. and GRINDLEY, N. D. F. (1984) Mutants of the $\gamma\delta$ resolvase: A genetic analysis of the recombination function. *Cell* **38**, 463-469.
- O'HARE, K. and RUBIN, G. M. (1983) Structures of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. *Cell* **34**, 25-35.
- PABO, C. O. and SAUER, R. T. (1984) Protein-DNA recognition. *Ann. Rev. Biochem.* **53**, 293-321.
- REED, R. R. (1981) Resolution of cointegrates between transposons $\delta\gamma$ and Tn3 defines the recombination site. *Proc. Natl. Acad. Sci. USA* **78**, 3423-3432.
- REED, R. R., SHIBUYA, G. I. and STEITZ, J. A. (1982) Nucleotide sequence of $\gamma\delta$ resolvase gene and demonstration that its gene product act as a repressor of transcription. *Nature* **300**, 381-383.
- SCHWARTZ, R. M. and DAYHOFF, M. O. (1978) *Atlas of Proteins Sequence and Structure*, Vol. 5 (ed. Dayhoff, M. O.), pp. 353-358. National Biomedical Research Foundation, Georgetown University Medical Center, Washington.