Biochem. J. (1989) 263, 19–23 (Printed in Great Britain)

Secondary structural features of the bacteriophage Mu-encoded A and B transposition proteins

George CHACONAS,* William D. McCUBBIN† and Cyril M. KAY†

*Department of Biochemistry, University of Western Ontario, London, Ontario N6A 5C1, Canada, and †Medical Research Council Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

The role of the bacteriophage Mu-encoded A and B proteins is to direct the transposition of Mu DNA. These are the first active DNA transposition proteins to have been purified and their mechanism of action at the biochemical level is under intensive study. Structural studies on these proteins, however, have lagged behind their biochemical characterization. We report here near- and far-u.v. c.d. spectra for these proteins and their secondary structural features derived from these data. The Mu A protein appears to be composed of primarily β-sheet (40%) with 24% α-helix, 9% β-turn and 27% random coil. In contrast, the Mu B protein contains 55% α-helix with only 13% β-sheet and 3% β-turn and 29% random coil. The near-u.v. c.d. spectrum of the A protein was not unusual; however, the profile of the B protein suggested either buried or restricted chromophores within the protein or short-range interactions between aromatic residues.

INTRODUCTION

In comparison with other DNA metabolic events, the biochemical study of DNA transposition has only been recently undertaken (see Howe & Berg, 1989, for reviews on DNA transposition). Because of the very low frequency of transposition, studies on this process were until recently limited to genetic approaches. The ability to increase the production of desired proteins by orders of magnitude using expression vectors was an essential step for the development *in vitro* of the first soluble transposition system (Mizuuchi, 1983) and subsequent biochemical studies. *In vitro*, this system utilizes the temperate bacteriophage Mu, a well-characterized bacterial transposon (see Pato, 1989, and Chaconas & Surette, 1988, for recent reviews). Study of the reaction *in vitro* has resulted in a wealth of previously unattainable information on the mechanism of Mu DNA transposition. In particular, the roles of the two Mu transposition proteins have been partially defined and studies on their structure and function have been initiated.

The Mu A protein is the pre-eminent transposition protein and is probably the strand transferase. The 75 kDa protein has been purified (Craigie & Mizuuchi, 1985*a*) and its primary structure derived from the sequence of the *A* gene (Harshey *et al.*, 1985). Binding of A *in vitro* at each of three sites at both the left and right ends of Mu DNA has been demonstrated (Craigie *et al.* 1984), and the importance of these binding sites for transposition *in vivo* has been corroborated (Groenen *et al.*, 1985; Groenen & van de Putte, 1986). The A protein can also bind to sequences in the early operator region about 1 kb from the left end of the Mu genome (Craigie *et al.*, 1984). Limited proteolysis has defined three domains within the A protein (Nakayama *et al.*, 1987). The *N*-terminal domain is approximately 30 kDa and exhibits site-specific binding to Mu DNA. The central core of 35 kDa binds DNA non-specifically and a temperature-sensitive transposition mutation has been mapped to this domain. The 10 kDa *C*-terminal domain has been suggested as a region for interaction with the Mu B protein (Harshey & Cuneo, 1986).

The Mu B protein is essential for efficient transposition. The 35 kDa protein has been purified (Chaconas *et al.*, 1985*b*) and its primary structure derived from the sequence of the Mu *B* gene (Miller *et al.*, 1984). The purified protein displays non-specific DNA-binding activity (Chaconas *et al.*, 1985*b*) and an ATPase activity which is stimulated by DNA and the Mu A protein (Maxwell *et al.*, 1987). The primary role of the protein appears to be the capture of a target DNA substrate for intermolecular transposition (Maxwell *et al.*, 1987; Surette & Chaconas, 1989; Craigie & Mizuuchi, 1987). The B protein is also the critical component for the phenomenon known as 'transposition immunity' (Adzuma & Mizuuchi, 1988). Partial proteolysis of Mu B has defined an *N*-terminal 25 kDa domain and an 8 kDa *C*-terminal domain (Teplow *et al.*, 1988). The *N*-terminal domain exhibits non-specific DNA binding and ATP binding (Teplow *et al.*, 1988). A region near the *N*-terminus shows sequence similarity to the helix–turn–helix DNA-binding motif of a number of DNA-binding proteins (Pabo & Sauer, 1984) and two internal regions within the protein show sequence similarity to the adenine nucleotide-binding folds of a variety of ATPases (Chaconas, 1987; Teplow *et al.*, 1988). Finally, a truncated form of B which lacks 18 amino acids from its *C*-terminus is functional for the conservative integration reaction, but not for replicative transposition, of Mu DNA (Chaconas *et al.*, 1985*a*). The *C*-terminal region of B may therefore be a site of interaction with the host replication machinery.

An intriguing feature of the Mu transposition reaction is the extensive network of protein–protein and protein–DNA interactions which mediate this process. The strand-transfer reaction, an early step in transposition, requires multiple DNA sites and at least three proteins: Mu A and B and *Escherichia coli* HU (Craigie *et al.*, 1985; Surette & Chaconas, 1989). The proteins and DNA sites assemble into stable complexes or trans-

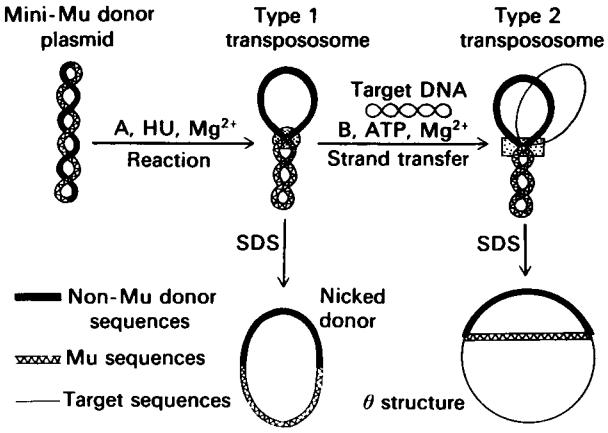

Fig. 1. The Mu DNA strand-transfer reaction

A type 1 complex, which is an intermediate in the strand-transfer reaction, is formed when a supercoiled mini-Mu donor plasmid is incubated with the Mu A protein and *E. coli* HU protein. The Mu ends are held together in a higher-order protein–DNA complex (type 1 transpososome) defining two topological domains: a relaxed non-Mu domain and a supercoiled Mu domain. Disruption of the complex with SDS results in the liberation of a nicked donor plasmid. The type 2 complex is the product of the strand-transfer reaction which remains complexed with protein (type 2 transpososome). In addition to type 1 reaction requirements, Mu B protein, ATP and target DNA are required for type 2 complex formation. The type 2 complex can be generated in a single reaction mixture, or by conversion of a preformed type 1 into a type 2 complex. Once a type 1 complex has been formed, supercoiling is not required for its conversion into the strand-transferred product. Disruption of the type 2 complex with SDS liberates the protein-free strand-transferred product or θ structure (Miller & Chaconas, 1986). Although the θ structure above has been drawn as a relaxed molecule for simplicity of presentation, it is important to note that the Mu DNA sequences, but not the vector or target DNA, are topologically constrained (Craigie & Mizuuchi, 1985*a*). Hence the Mu DNA stem retains the Mu supercoils originally present in the type 2 complex. This Figure is from Surette & Chaconas (1989) and was adapted from Surette *et al.* (1987). Reproduced with permission from the American Society for Biochemistry and Molecular Biology.

posomes (see Fig. 1), which mediate the reaction (Surette & Chaconas, 1989). The structure of these complexes remains to be elucidated and will be largely dependent upon the three-dimensional structures of the individual proteins. Although the structure of HU has been determined by X-ray crystallography, the preparation of diffractable crystals has not been reported for any transposition protein. As a step towards understanding the structure and function of the Mu A and B proteins we have analysed their secondary structure using c.d.

EXPERIMENTAL PROCEDURES

Protein purification

The Mu A protein was purified by a modification of the procedure of Craigie & Mizuuchi (1985*b*) as previously described (Surette *et al.*, 1987). The Mu B protein was purified by our published protocol (Chaconas *et al.*, 1985*b*). Both proteins were active in the Mu DNA strand-transfer reaction *in vitro* as described by Surette *et al.* (1987). The concentration of the A protein was determined by amino acid analysis of an A protein hydrolysate using a norleucine standard. B protein concentration was determined using the absorption coefficient $A_{280}^{1\ mg/ml} = 1.16$ (Chaconas *et al.*, 1985*b*).

C.d. measurements

These measurements were made on a Jasco 500C spectropolarimeter interfaced with a DP 500N data processor under constant nitrogen flushing. The cell was maintained at 25 °C using a Lauda K-2/R circulating water bath. Near-u.v. (320–250 nm) scans were performed in a microcell with a path length of 1 cm which required only 90 μl of solution. Far-u.v. (250–190 nm) runs were done in a cell with a 0.0102 cm path length. Protein concentrations were as indicated in the Figure legends. The computer-averaged trace of 16 scans was employed in all calculations. Signal due to solvent was subtracted. The instrument was routinely standardized with D-(+)-10-camphorsulphonic acid and pantoyl-lactone according to procedures outlined by the manufacturers. The data were normally plotted as mean-residue-weight ellipticity ($[\theta]$; units of degrees $\cdot$ cm^2 $\cdot$ dmol^{-1}) versus wavelength in nm. The mean residue weight was taken to be 115.

Fluorescence measurements

Fluorescence emission and excitation spectra were measured on a Perkin–Elmer MPF 44B recording spectrofluorimeter equipped with the DCSU-2 corrected spectra accessory which allows for automatic subtraction of fluorescence due to solvent. The instrument was operated in the ratio mode, and cells were thermostatically maintained at 20 °C. The emission and excitation slits were set at 5 nm and a 1 cm square cell was used.

Computer-assisted analyses

In addition to plotting the far-u.v. c.d. data in terms of mean residue ellipticity versus wavelength, the secondary structure was determined by using the mainframe-drive FORTRAN program CONTIN, developed by Provencher & Glöckner (1981), which analyses c.d. spectra as the sum of data collected from 16 proteins the structures of which are known from X-ray crystallography. The input to the program was the mean residue ellipticities in 1 nm intervals from the minimum value measured to 240 nm. A subroutine built into the CONTIN program takes into consideration errors in estimation of protein concentration of $\pm 3\%$. Within these limits the error in estimating the secondary structure content is of the order of $\pm 0.5\%$. Predictions for the secondary structure were based upon the Dufton & Hider (1977) modification of the Chou & Fasman (1974) method. In addition, for the Mu A protein, predictions were also made by the procedure of Cid *et al.* (1982), in which the secondary structure parameters are extracted from basic hydrophobicity and accessibility profiles.

A composite surface profile was drawn showing the regions of sequence most likely to lie on the surface of the protein. This computer-generated profile is a composite of three approaches, which include h.p.l.c., hydrophilicity parameters (Parker *et al.*, 1986), accessibility (Janin, 1979) and flexibility (Karplus & Schulz, 1985).

RESULTS AND DISCUSSION

Secondary structural features of the Mu A and B proteins

The generation of diffractable crystals of the Mu A and B proteins has not yet been reported. Because of the propensity of these proteins to aggregate (G. Chaconas, unpublished work), X-ray diffraction may not be a viable option for structural studies. As a first step in characterizing the structure of these proteins, we analysed their far-u.v. c.d. spectra. The spectra of both proteins are shown in Fig. 2. The data were analysed for secondary structure by the computer program CONTIN by Provencher & Glöckner (1981). The A protein displayed a definite shoulder near 220 nm and a pronounced minimum near 203 nm, signifying only a moderate amount of α-helix. Computer analysis revealed 28% α-helix, 29% β-sheet and 12% β-turn with 36% aperiodic form at 25 °C. Because of the relatively high proportion of random coil at 25 °C, a spectrum at 6 °C was also obtained. This spectrum was similar to the one at higher temperature and quantitative analysis of the data revealed 24% α-helix, 40% β-sheet, 9% β-turn and 27% random coil. Thus the protein possesses a moderate amount of α-helix, but the main structural elements seem to be β-sheet/turns. These results are not in good agreement with the secondary structural features predicted from the primary sequence by the method of Chou & Fasman (1974), which are 46% α-helix, 1% β-sheet, 32% β-turn and 4% random coil. This discrepancy could be due to the amount of helical structure in the protein being reduced because of aggregation or reduced thermal stability; both features are quite common to proteins having extensive amounts of β-structure, (McCubbin *et al.*, 1988). Predictions based on the method of Cid *et al.* (1982) yielded 31% α-helix, 24% β-sheet, 14% β-turn and 31% aperiodic structure. These values are much closer to the experimentally observed ones. The composite surface values based upon hydrophilicity, accessibility and flexibility parameters (Parker *et al.*, 1986) are shown in Fig. 3.

The Mu B protein displayed a bimodal minimum at 222 nm and 208 nm, which is typical for a helical protein. Analysis of the c.d. data revealed 55% α-helix, 13% β-sheet, 3% β-turn and 29% aperiodic structure. These results are in good agreement with the predicted results from analysis of the primary structure, which estimated 58% α-helix, 14.7% β-sheet, 13.5% β-turn and 13.8% random coil. In comparing the observed and predicted values it is important to note that the CONTIN program only analyses for one type of β-turn, while the Chou–

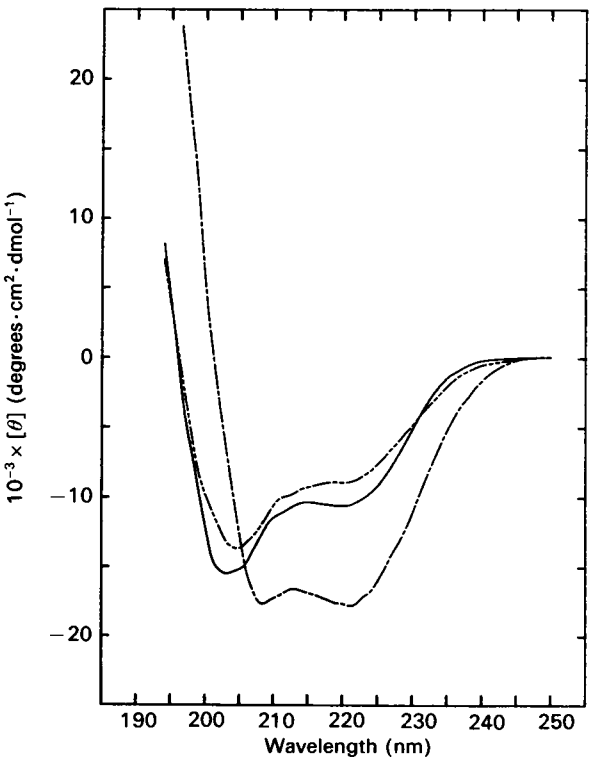

Fig. 2. Far-u.v. c.d. spectra of the bacteriophage Mu-encoded A and B transposition proteins

The solvent for the A protein was 25 mM-Hepes/NaOH, pH 7.6, 0.6 M-KCl, 0.1 mM-EDTA, 0.3 mM-dithiothreitol, 10% (v/v) glycerol, and the protein concentration was 0.19 mg/ml. The B protein was analysed at 25 °C in 30 mM-Tris/HCl, pH 7.8, 1.0 M-NaCl, 1 mM-EDTA, 10 mM-β-mercaptoethanol, and the protein concentration was 0.56 mg/ml. A protein at 6.1 °C (——) and 25 °C (- - - -); B protein (— - —).

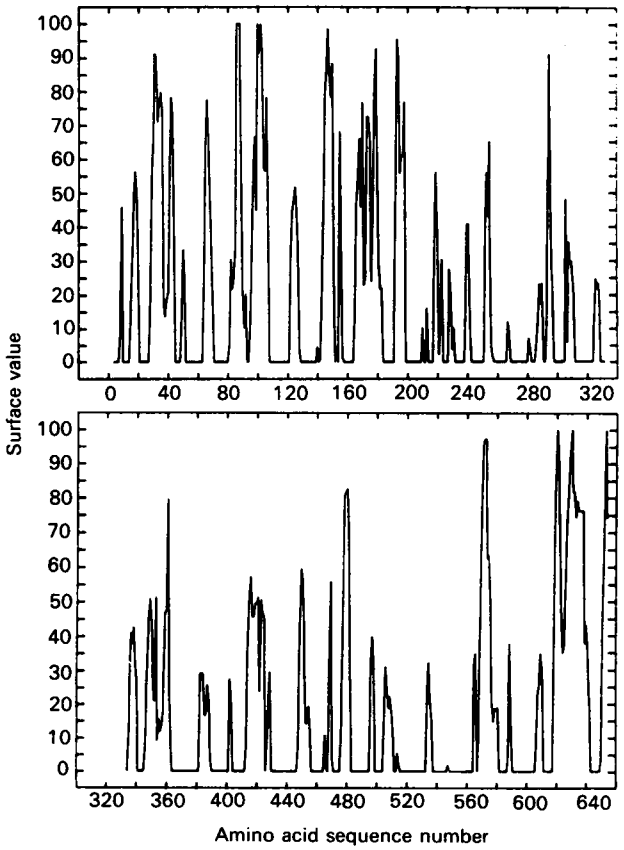

Fig. 3. Graphical display of composite plot of hydrophilicity, surface accessibility and polypeptide flexibility (Parker *et al.*, 1986) for the Mu A protein

Our computer program required that the sequence data be employed in two stages, the *N*-terminal half, residues 1–330 and residues 331–662. The portions of the profile above the '50'-line are predicted to lie on the surface of the A protein.

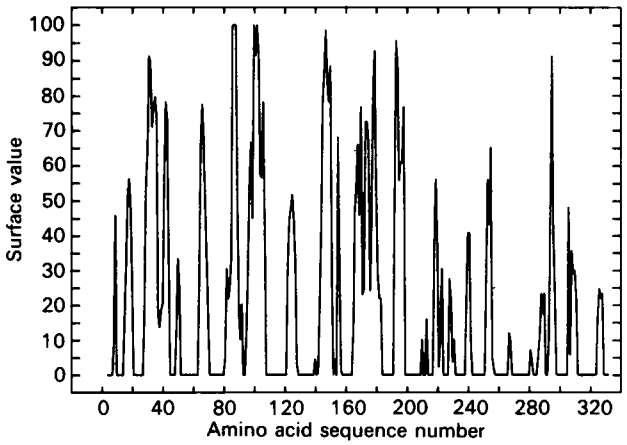

Fig. 4. Graphical display of composite plot of hydrophilicity, surface accessibility and polypeptide flexibility (Parker *et al.*, 1986) for the Mu B protein

The portions of the profile above the '50'-line are predicted to lie on the surface of the B protein.

Fasman prediction includes all the β-turns and β-bends or β-bulges in the protein. Thus a more proper comparison between the observed and predicted secondary structure should combine the β-turn and random-coil outputs of each program. When these terms are summed, the observed β-turn and random coil comprise 32% and the predicted values are in good agreement at 27.3%. The predicted surface values are shown in Fig. 4. The good correspondence between the observed and predicted secondary structure is important for the Mu B protein. Because of the insolubility of this protein at low ionic strength, the only purification protocol requires several chromatographic steps in 6.2 M-urea (Chaconas *et al.*, 1985*b*). Although the renatured protein is fully active in the transposition reaction, the far-u.v. c.d. data support the contention that most of the B molecules have properly refolded into their native conformation.

Analysis of aromatic residues

The near-u.v. c.d. spectrum of the A protein was not unusual (Fig. 5). At 6 °C the near-u.v. c.d. spectrum was dominated by a negative trough centred at 274 nm, $[\theta]_{274} = -86°$, probably representing contributions from Tyr residues in asymmetric environments. Fine structure, on this limb from 250 nm to 270 nm was probably due to Phe contributions. A small positive peak centred near 294 nm was perhaps due to Trp. The spectrum noted at 25 °C was quite similar to that seen at low temperature; however, there seemed to be more fine structure and the minimum was much broader (Fig. 5).

The near-u.v. c.d. profile of the B protein (Fig. 5) is rather interesting. The spectrum was dominated by a large trough having a minimum near 276 nm with a value of $-125°$. This figure was larger than expected from the Trp/Tyr content of the protein (4 and 5 residues, respectively). The large negative value may be related to the presence of buried/restricted chromophoric residues in the molecule. Another good possibility, especially in light of the results of radioiodination experiments (G. Chaconas, unpublished work), which suggest complete accessibility of all Tyr residues, is that some of the aromatic residues may be in close proximity and involved

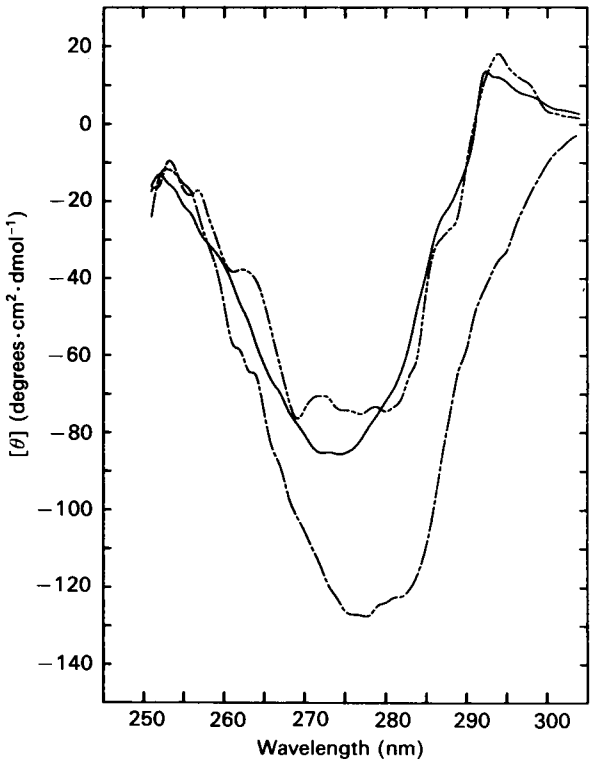

Fig. 5. Near-u.v. c.d. spectra of the Mu A and B proteins

The solvents and symbols used are described in the legend to Fig. 2.

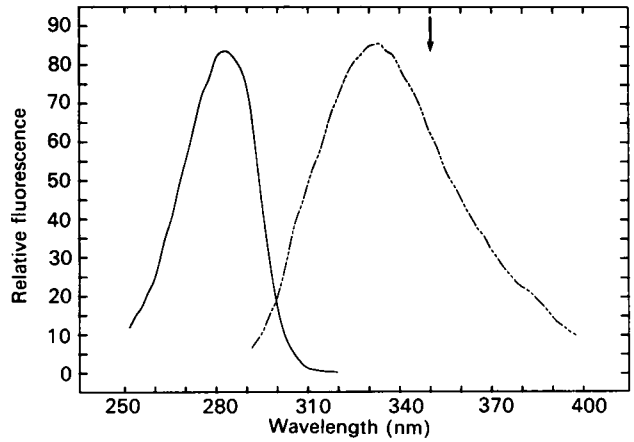

Fig. 6. Fluorescence excitation and emission spectra of the Mu B protein

The solvent used is described in the legend to Fig. 2 and the protein concentration was 5.6 μg/ml. The arrow at 350 nm indicates the emission maximum noted for fully exposed Trp residues. Excitation spectrum (——); emission spectrum (- - - -).

in short-range interactions. A similar phenomenon has been noted with the highly helical muscle protein tropomyosin, where Tyr–Tyr interactions at distances of less than 8 Å are believed to occur between aligned helical chains (Bullard *et al.*, 1976). To investigate further the possible presence of solvent-inaccessible aromatic residues, two additional experiments were performed. (1) An

alkaline titration of the B protein followed by near-u.v. c.d. analysis revealed the gradual disappearance of the minimum peak at 275 nm as the pH increased; the 240–250 nm region also became more positive (results not shown). These trends are usually associated with the ionization of Tyr. (2) Fluorescence emission and excitation spectra were obtained as shown in Fig. 6, and are typical of Trp-containing proteins. The emission maximum is near 333 nm, which suggests burying of the fluorophores within the protein structure.

Consideration of the composite surface plot for Mu B protein (Fig. 4) indicates that the five Tyr residues are either in, or juxtaposed to, exposed surface regions (this explains the accessibility of these residues to iodination), whereas three of the four Trp residues are included in elements of α-helix or β-structure and are unlikely to be surface exposed. Trp-55, however, is in a short segment of sequence which may lie on the surface. Further studies will be required to elucidate the unusual structural arrangement of the aromatic residues in the B protein which result in the near-u.v. c.d. spectrum reported here.

This work was supported by grants from the Medical Research Council of Canada to G.C. and C.M.K. G.C. is the recipient of a Scientist Award from the Medical Research Council of Canada.

REFERENCES

Adzuma, K. & Mizuuchi, K. (1988) Cell (Cambridge, Mass.) **53**, 257–266

Bullard, B., Mercola, D. A. & Mommaerts, W. F. H. M. (1976) Biochim. Biophys. Acta **435**, 90–99

Chaconas, G. (1987) in Phage Mu (Symonds, N., Toussaint, A., van de Putte, P. & Howe, M. M., eds.), pp. 137–157, Cold Spring Harbor Laboratory, Cold Spring Harbor

Chaconas, G. & Surette, M. (1988) BioEssays **9**, 205–208

Chaconas, G., Giddens, E. B., Miller, J. L. & Gloor, G. (1985*a*) Cell (Cambridge, Mass.) **41**, 857–865

Chaconas, G., Gloor, G. & Miller, J. L. (1985*b*) J. Biol. Chem. **260**, 2662–2669

Chou, P. Y. & Fasman, G. D. (1974) Biochemistry **13**, 222–245

Cid, H., Bunster, M., Arriagada, E. & Campos, M. (1982) FEBS Lett. **150**, 247–254

Craigie, R. & Mizuuchi, K. (1985*a*) Cell (Cambridge, Mass.) **41**, 867–876

Craigie, R. & Mizuuchi, K. (1985*b*) J. Biol. Chem. **260**, 1832–1835

Craigie, R. & Mizuuchi, K. (1987) Cell (Cambridge, Mass.) **51**, 493–501

Craigie, R., Mizuuchi, M. & Mizuuchi, K. (1984) Cell (Cambridge, Mass.) **39**, 387–394

Craigie, R., Arndt-Jovin, D. J. & Mizuuchi, K. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 7570–7574

Dufton, P. Y & Hider, R. C. (1977) J. Mol. Biol. **115**, 177–193

Groenen, M. A. M. & van de Putte, P. (1986) J. Mol. Biol. **189**, 597–602

Groenen, M. A. M., Timmers, E. & van de Putte, P. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 2087–2091

Harshey, R. M. & Cuneo, S. D. (1986) J. Genet. **65**, 159–174

Harshey, R. M., Betzoff, E. D., Baldwin, D. L., Miller, J. L. & Chaconas, G. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 7676–7680

Howe, M. M. & Berg, D. (eds.) (1989) Mobile DNA, American Society for Microbiology, Washington, in the press

Janin, J. (1979) Nature (London) **277**, 491–492

Karplus, P. A. & Schulz, B. E. (1985) Naturwissenschaften **72**, S 212

Maxwell, A., Craigie, R. & Mizuuchi, K. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 699–703

McCubbin W. D., Kay, C. M., Narindnasorasak, S. & Kisilevsky, R. (1988) Biochem. J. **256**, 775–785

Miller, J. L. & Chaconas, G. (1986) Gene **48**, 101–108

Miller, J. L., Anderson, S. K., Fujita, D. J., Chaconas, G., Baldwin, D. L. & Harshey, R. M. (1984) Nucleic Acids Res. **12**, 8627–8638

Mizuuchi, K. (1983) Cell (Cambridge, Mass.) **35**, 785–794

Nakayama, C., Teplow, D. B. & Harshey, R. M. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 1809–1813

Pabo, C. O. & Sauer, R. T. (1984) Annu. Rev. Biochem. **53**, 293–321

Parker, J. M. R., Guo, D. & Hodges, R. S. (1986) Biochemistry **25**, 5425–5432

Pato, M. L. (1989) in Mobile DNA (Howe, M. M. & Berg, P., eds.), American Society for Microbiology, Washington, in the press

Provencher, W. W. & Glöckner, J. (1981) Biochemistry **20**, 33–37

Surette, M. G. & Chaconas, G. (1989) J. Biol. Chem. **264**, 2028–2034

Surette, M. G., Buch, S. & Chaconas, G. (1987) Cell (Cambridge, Mass.) **49**, 253–263

Teplow, D. B., Nakayama, C., Leung, P. C. & Harshey, R. M. (1988) J. Biol. Chem. **163**, 10851–10857

Received 10 February 1989/17 April 1989; accepted 24 April 1989