Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 7676–7680, November 1985
Genetics

Primary structure of phage Mu transposase: Homology to Mu repressor

(transposable elements/phage D108/integration host factor/DNA-binding protein/secondary structure prediction)

RASIKA M. HARSHEY*, ELIZABETH D. GETZOFF*, DONALD L. BALDWIN*, JANET L. MILLER†, AND GEORGE CHACONAS†

*Department of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, CA 92037; and †Cancer Research Laboratory and Departments of Biochemistry and Microbiology-Immunology, University of Western Ontario, London, ON N6A 5B7 Canada

Communicated by Franklin W. Stahl, July 19, 1985

ABSTRACT **The phage Mu transposase is essential for integration, replication–transposition, and excision of Mu DNA. We present the complete nucleotide and derived amino acid sequence of the transposase and analyze implications for transposase/DNA interaction. The NH_2 terminus of the Mu transposase has considerable sequence homology with the Mu repressor and with the NH_2 terminus of the transposase of the Mu-like phage D108. These three proteins are known to share binding sites on DNA. The protein sequence and predicted secondary structural similarities at the NH_2 termini of the three proteins suggest a common DNA-binding region similar to the regions found in proteins of known structure. An internal sequence in the Mu A protein also shares these features. We anticipate that these regions will be involved in DNA recognition during transposition.**

The temperate phage Mu is remarkably efficient at transposing its DNA into multiple sites in many bacterial genomes and mediating a variety of DNA rearrangements (1, 2). Mu is therefore an excellent model for studying protein/DNA interactions involved in transposition and in associated chromosome shuffling. Transposition requires two phage-encoded proteins: the transposase (encoded by gene *A*) and the transposition enhancer (encoded by gene *B*). Unlike other transposons, Mu has dissimilar sequences at its left and right ends (3). However, A protein apparently binds three specific blocks of sequences at each end of the DNA, allowing identification of a consensus sequence recognized by A protein (4). In addition to binding of Mu ends, transposition requires binding to target DNA and appropriate cutting and strand transfer reactions.

The *A* gene extends from 1.3 to 3.3 kilobases (kb) from the left end of Mu and encodes a 70-kilodalton protein (5), which has been purified (6). Expression of the early genes of Mu, including *A*, is regulated by the repressor c, which binds to an operator sequence and shuts off early transcription (7). The repressor c, at high concentrations, can occupy almost exactly the same sites on Mu ends as the A protein does, and conversely, A can bind to fragments containing the Mu operator sequence (4). This implies that A and c are related and may interact in the control of transposition.

Also related to A is the transposase of the Mu-like phage D108 (8). Electron-microscopic analysis of Mu–D108 heteroduplexes shows that, except for three small regions, the DNAs of the two phages are homologous. One nonhomologous area extends across the repressor gene into the NH_2 terminus of the *A* gene and includes the operator sequence bound by the repressor. Accordingly, Mu and D108 have different immunities (i.e., their repressors do not bind each other's operators). Demonstrated sequence differences at the NH_2 termini of the two *A* genes support the electron-microscopic analysis (9). However, the A proteins from the two phages can function interchangeably to promote transposition, although with different efficiencies, and Mu A binds at the left end of D108 to DNA sequences similar to those at Mu ends (4). Thus Mu and D108 A proteins appear to share DNA-binding specificity for transposition but probably not for operator recognition.

Previous studies of Mu defined 264 nucleotides at the NH_2 terminus of the *A* gene (10). To understand DNA recognition during transposition, we have sequenced the entire *A* gene and identified sites in the transposase that may govern transposition. We identify homologous regions of Mu transposase, Mu repressor c, Mu transposition-enhancer B, and phage D108 transposase that resemble the α-helix–turn–α-helix structure implicated in many sequence-specific DNA-binding proteins (11, 12).

MATERIALS AND METHODS

Bacterial and Phage Strains and Plasmids. Mu DNA fragments from plasmids pCL222 (13), pRA600 and pGC511 (14, 15), and pTM211 (16), were subcloned in M13 phage vectors and sequenced. Phages M13mp8 and mp9 and their host strain *Escherichia coli* JM103 were obtained from Bethesda Research Laboratories and were propagated as described by this supplier's manual.

DNA Sequencing Strategy. Sequencing reactions were carried out with a modified version of the dideoxynucleotide chain termination method (17) as described in the Bethesda Research Laboratories sequencing manual. The DNA fragments used for cloning in M13 and sequencing are identified in Fig. 1. The sequence of Mu transposase was derived from three regions of the gene (Fig. 1*B*) and the two segments overlapping them as follows: (*i*) *Bal* I–*Pst* I fragment of pCL222 was isolated from agarose gels, made blunt ended with T4 DNA polymerase, and inserted into the *Sma* I site of M13mp8. Clones with both orientations of the insert were isolated and sequenced. (*ii*) The *Pst* I–*Bgl* I fragment from pRA600 was isolated, ligated to itself, and sheared by sonication to generate subfragments of average size 200–600 base pairs (bp). These were made blunt ended with T4 DNA polymerase, and fragments in the size range 300–600 bp were isolated by trough elution employing preparative agarose gels as described (18). The mixture of random subfragments was inserted into the *Sma* I site of M13mp9. Individual clones were isolated and sequenced on both strands. On an average, each base pair was sequenced six times. The random sequence was compiled into a contiguous stretch by using the Staden computer program (19). (*iii*) The *Bgl* I fragment from

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: kb, kilobase(s); bp, base pair(s); CAP, catabolite gene activator protein.

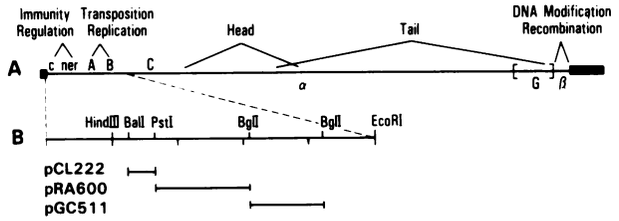

FIG. 1. (*A*) Schematic map of the Mu genome (not to scale). The 37-kb DNA is flanked by host sequences (shaded bars). The early region contains the immunity determinant *c*, another negative regulator *ner*, and the primary early genes involved in DNA transposition, *A* and *B*. Between genes *c* and *ner* lie the operator region bound by the repressor c and the promoters for the repressor (transcribed leftward) and the early genes (transcribed rightward). (*B*) Restriction map of the early region. The DNA length scale is indicated by short vertical lines placed at 1-kb intervals below the horizontal line of the map. Below are the DNA fragments from the corresponding plasmids used for cloning in M13 and sequencing.

pGC511 was isolated and inserted into the *Sma* I site of M13mp8. Subclones of this fragment generated with *Dde* I, *Alu* I, *Rsa* I, *Hae* III, and *Dra* I and inserted into the *Sma* I site of M13mp8 were also sequenced. Both strands of the DNA were thus sequenced. (*iv*) The sequence across the two joints at *Pst* I and *Bgl* I was confirmed as follows: A 380-bp *Hin*fI fragment from pTM211 that includes the *Pst* I site and a 160-bp *Ban* I–*Hin*fI fragment that includes the *Bgl* I site were isolated. Each was subcloned in the *Sma* I site of M13mp9 and sequenced in both orientations.

Protein Purification and Sequencing. The Mu A protein was purified according to the published procedure (6). A confirming partial NH_2-terminal amino acid sequence was determined by R. Aerbersold, D. Teplow, and S. Kent at the California Institute of Technology (personal communication).

Sequence Analysis. The simultaneous alignment of multiple protein sequences (done by hand) emphasized the alignment of both identical and similar residues to yield maximal homology. Similarities considered included like hydrophobicity, size, charge, or secondary structure preferences. Computer programs from the Protein Sequence Database of the Protein Identification Resource‡ were used to align pairs of sequences, to analyze homologies versus random comparisons, and to search for DNA homologies within the transposase gene. Protein secondary structure predictions and hydrophobicity profiles were calculated by the computer program SECSTR (unpublished method), which applied Chou/Fasman empirical parameters (20) and the Eisenberg consensus hydrophobicity scale (21).

RESULTS AND DISCUSSION

Primary Structure. The three DNA fragments used for cloning and sequencing gene *A* are aligned on a diagram of the Mu genome in Fig. 1. The longest open reading frame (Fig. 2) extends from position 1328 to 3316 and encodes a protein of 662 amino acids with a predicted molecular weight of 74,889. The amino acid sequence of the first 8 residues of the purified protein agrees with that deduced from the DNA sequence.

Possible Repressor and Transposase Binding Sites Overlap Gene *A*. Besides binding to the Mu ends and operator region, the Mu A and repressor proteins bind *in vitro* to sequences within the coding region of *A* (4); thus the two proteins may interact in additional ways to regulate transposition. Therefore, we looked within the *A* coding region of the DNA for sequences similar to the consensus sequence defined from the three stronger transposase-binding sites at Mu ends (4): TGNTTCANTNNAARYRCGAAAR. The best match (four mismatched bases) occurs at nucleotides 2849–2870 (Fig. 2). For one of the three strong transposase binding sites, the right half of this consensus sequence can be deleted without affecting transposition (24). Interestingly, the left end of the consensus sequence—TGNTTCANT—resembles subsets of the operator sequence recognized by the repressor (7), suggesting that both proteins may recognize this sequence. The DNA sequence starting at position 2812 matches this repressor recognition sequence fairly well and is also similar to the extended consensus sequence for the transposase binding site. These or other similar sites may account for the *in vitro* binding of the two proteins to the *A* gene.

Homology with Mu c, D108 A, and DNA-Binding Proteins of Known Structure. We have identified significant sequence homology among the NH_2-terminal regions of Mu A, Mu c, and D108 A proteins. Amino acids common to all three sequences cluster in the region numbered 42–57 (Fig. 3), which also includes the region of maximal homology between the Mu transposase and repressor. Significant homology between these two proteins also occurs between positions 109 and 149 (Fig. 3). In contrast, the best homology between the two transposases starts from position 65, suggesting that this region may be specific to the transposition function. Mu c and D108 A share less sequence homology with each other than either does with Mu A. The homology among these three sequences may reflect their shared ability to bind Mu ends.

The three site-specific DNA-binding proteins for which three-dimensional structures are known [the cro and cI repressors of phage λ (25, 26), and the catabolite gene activator protein (CAP) of *E. coli* (27)], each contain an α-helix–turn–α-helix structural motif thought to be responsible for DNA binding. We determined and plotted secondary structure predictions and hydrophobicity profiles for the Mu transposase, Mu repressor, and D108 transposase protein sequences (Fig. 4). The overall match of each set of superimposed curves suggests that these three proteins have similar secondary structures for the region corresponding to the first two-thirds of the repressor sequence. The predictions indicate four α-helical peaks near the NH_2 terminus, each separated from the next by a tight turn. This α-helix–turn predicted region could be viewed as three overlapping α-helix–turn–α-helix structural motifs (as labeled 1, 2, and 3 in Fig. 4*A*), resembling those found and predicted for DNA-binding proteins. The similar hydrophobicity profiles through position 145 for the Mu A and c and D108 A proteins (Fig. 4*C*) are consistent with their primary and predicted secondary structural similarities.

Sequence homologies have been identified in DNA-binding regions of proteins with known or proposed bihelical structures (Fig. 5; refs. 11, 12, and 28). Gly is preferred at position 9 as part of the tight turn, and the side chains at positions 5 (Ala or Gly preferred) and 15 (Val or Ile preferred) form van der Waals contacts, which probably help to maintain the proper angle between the two α-helices. Examination of the protein sequences for Mu and D108 transposases and Mu repressor in the potential α-helix–turn–α-helix regions of the sequence (Figs. 3 and 4) shows that the middle bihelical region best fits this sequence pattern (Fig. 5). For this bihelical region, positions 5 and 9 would both be Gly in all three proteins. Allowing one insertion before residue 14, position 15 would be Val, Ala, and Ile in Mu c, Mu A, and D108 A, respectively. Stereo illustrations of the α-helix–turn–α-helix structures of CAP, cro, and cI (26, 27, 29) suggest that such an insertion could be accommodated within

‡Barker, W. C., Chen, H. R., Hunt, L. T., Orcutt, B. C., Yeh, L. S., George, D. G., Johnson, G. C., Seibel-Ross, E. I. & Dayhoff, M. O. (1984) Nucleic Acid Sequence Database (National Biomedical Research Foundation, Georgetown University Medical Center, Washington, DC).

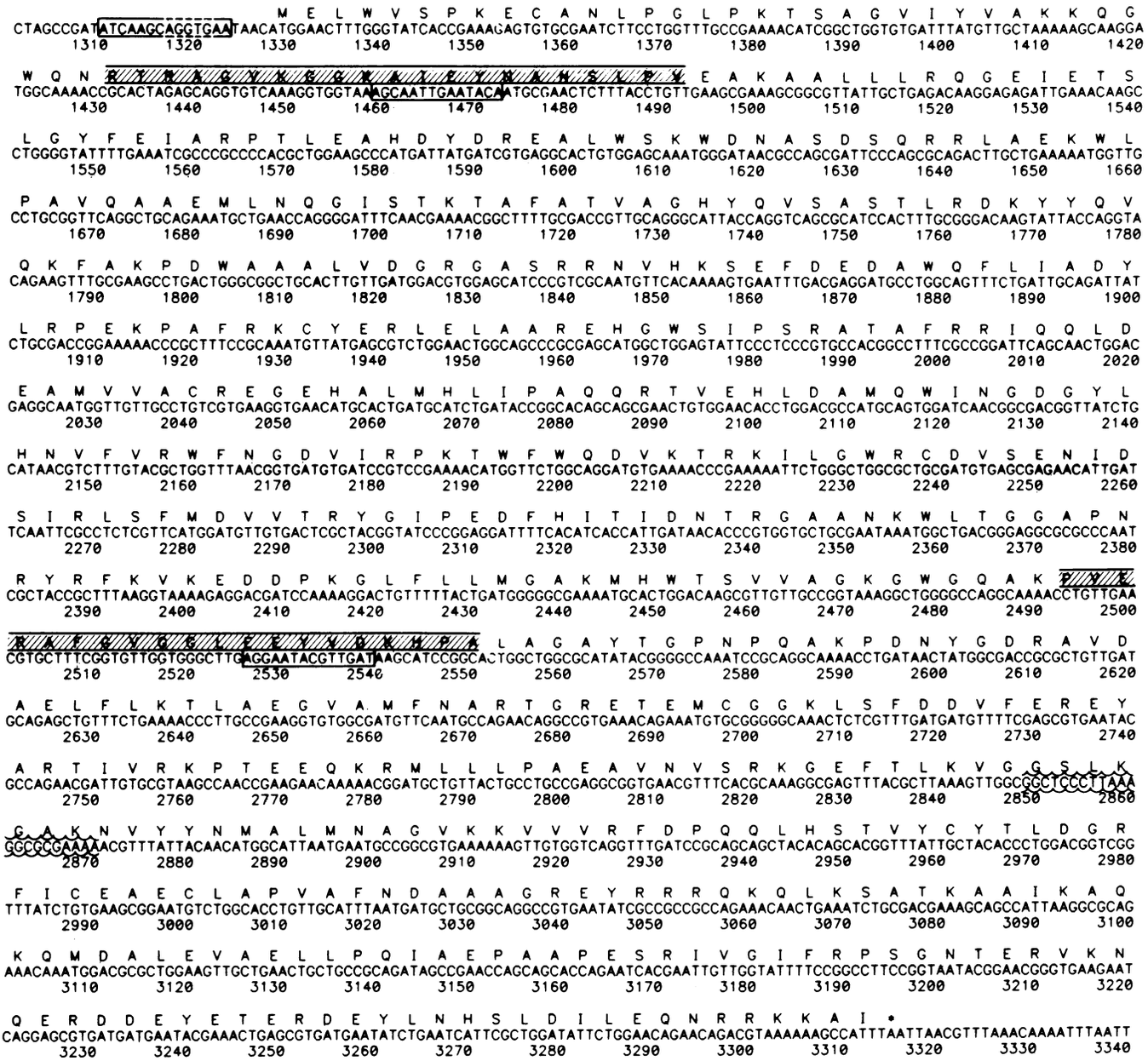

FIG. 2. Nucleotide and predicted amino acid sequence for Mu *A*. The sequence is shown 5′ to 3′, beginning 1300 bp from the left end of the Mu genome and extending to 3340. The standard one-letter code for amino acid residues is used. Broken lines indicate the Shine–Dalgarno sequence AGGTG. Wavy lines mark similarity to the A protein recognition sequence at Mu DNA ends. The termination codon is indicated by an asterisk. Sequences similar (two mismatches) to the integration host factor (IHF) consensus binding site TNYAANNNRTTGAT (22) (Y = pyrimidine, R = purine) are boxed. Those at 1310 and 1459 occur on the opposite strand in the inverse orientation. Mutations in *himA* and *himD* genes, whose products constitute IHF, affect early Mu transcription (7) and block transposition (23). The hatched amino acid sequences (residues 35–55 and 389–408) resemble the α-helix–turn–α-helix structure present in a number of DNA-binding proteins (see Fig. 5).

the turn or adjoining helical ends without requiring a concomitant shift of critical side-chain positions within the helices. Similar insertions are tolerated in turns between helices found in related pairs of proteins [e.g. between different species of hemoglobin (30) or cytochrome *c*′ (31)]. Furthermore, the residues aligned at positions 4, 5, 8, 15, 18, and 19 (which face the hydrophobic core in the known protein structures) are, in general, nonpolar in the Mu and D108 sequences, as is true in other DNA-binding proteins (12). A different potential α-helix–turn–α-helix sequence (positions 130–150 in Fig. 3) was identified in Mu A and Mu c by R. Brennan and B. Matthews (personal communication). They compared amino acid sequences and minimum base changes from a master set of 10 DNA-binding proteins but did not allow insertions and deletions. Their site is favorable in that it requires no insertions, but unfavorable in replacing the preferred Ala (position 5 in Fig. 5) with the much bulkier Tyr.

The bihelical motif in the three proteins, identified on the basis of sequence similarities to DNA-binding proteins (Fig. 5), shows reasonably good agreement with that predicted from secondary structure considerations (Fig. 4*A*: the middle pair of α-helices). In the Mu repressor and Mu and D108 transposases, the first α-helical region marked in Fig. 4 could correspond to the additional α-helix found in CAP, cro, and cI, which precedes the bihelical unit and has only limited structural homology among those three proteins.

The transposition protein B from Mu is required for replicative transposition and is responsible for the high transposition frequency. A sequence in Mu B (residues 21–40; ref. 14), similar to known DNA-binding sequences, is

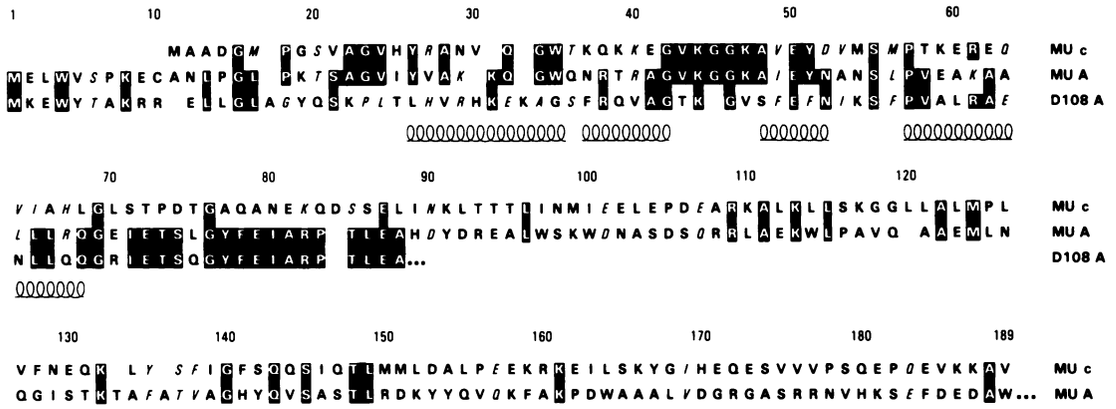

FIG. 3. Alignment of the Mu A, Mu c, and D108 A amino acid sequences. Numbering is based upon the total alignment. Sequence-invariant residues are shown on a black background, and chemically similar amino acids are indicated by italic letters. The sets within which residues were considered similar are: (F, W, Y), (F, I, L, M, V), (A, G), (G, P), (D, N), (Q, N), (E, Q), (D, E), (H, K, R), and (S, T). Computer alignments[‡] of Mu transposase with Mu repressor and D108 transposase indicate significant homology (10 standard deviations above the mean score for randomly matched sequences). The regions of predicted α-helical secondary structure are indicated schematically below the sequences. Predicted turns occur between the noted helical areas, and both turns and helices clearly fall within the major regions of sequence similarity. The turn predicted between the last two helices would intervene in the second half of the bihelical motif shown in Fig. 5, but it could form a 3_{10} helix. A similar intervening turn is predicted for the sequence of phage λ cI repressor, but it is not present in the structure.

aligned in Fig. 5 and was also selected by R. Brennan and B. Matthews (see above). Our secondary structure predictions for Mu B (not shown) identify an α-helix–turn motif but no putative second helix. However, the second helix is also poorly predicted in our analysis of CAP or cI repressor sequences. Moreover, there are striking sequence similarities between this protein B sequence and the aligned sequence (residues 36–56) in D108 transposase (Fig. 5). Another

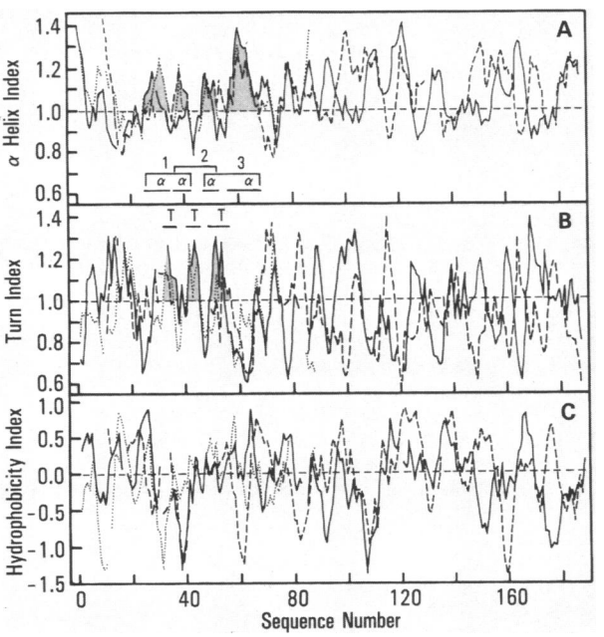

FIG. 4. Superimposed profiles of secondary structure and hydrophobicity for the NH_2-terminal region of the Mu A sequence (solid line), the complete sequence of Mu c (dashed line), and a partial sequence of D108 A (dotted line). Breaks indicate deletions and numbering is based on the combined alignment. (*A*) α-Helix conformational prediction values (20) averaged over five residues and plotted at the central residue. Regions of probable α-helix occur above the indicated cutoff of 1.0. Bars at the bottom represent the four likely helical regions (the shaded peaks) common to the three proteins, which occur in regions of sequence similarity with DNA-binding proteins. (*B*) Predicted turn conformation (20) averaged over four residues and plotted at the first residue. The three predicted turns (shaded peaks) common to these proteins in regions of similarity with DNA-binding proteins are marked by bars above the curve. (*C*) Comparison of the hydrophobicity profiles (21) as averaged over five residues and plotted at the central residue. Overall patterns of hydrophobicity are well matched for the first two-thirds of the aligned sequences (residues 1–145 on this curve) but not for COOH-terminal regions.

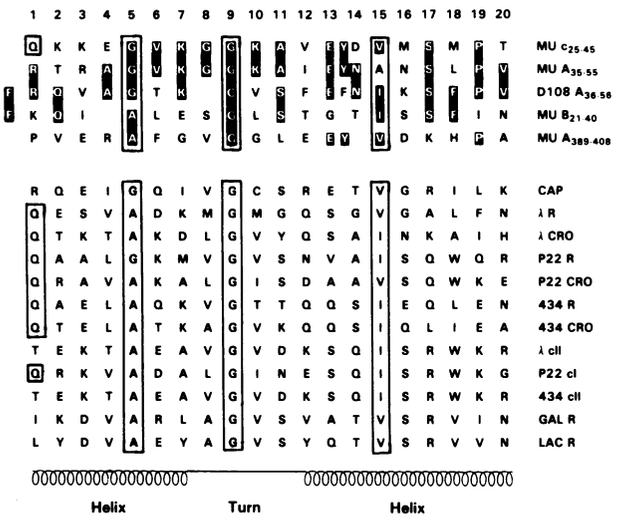

FIG. 5. Alignment of Mu c, A, B, and D108 A sequences with those of known DNA-binding proteins (12, 28). Simultaneous alignment of multiple sequences allows identification of similarities not statistically significant in pairwise alignments. Critical conserved residues are boxed, and the α-helix–turn–α-helix structural motif is aligned at the bottom. In the Mu and D108 protein sequences, identical residues are shown in white on a black background and sequence numbers are given at the far right. Using a similar group of 20-residue segments from 11 repressors, Sauer *et al.* (28) found 43–57 identities when each was compared with all of the remaining set. An equivalent size sample from the above alignment (the top 11 sequences) has 32–46 identities (random expectation = 10.8), with the best match being for D108 A. Together with the similarities in their predicted secondary structure, this level of homology is certainly significant. Specific sequence similarities between Mu B and D108 A within this predicted motif include: Phe followed consecutively by a positively charged amino acid, Gln, and a branched hydrophobic side chain in the first helix; Gly, a branched hydrophobic side chain, and Ser in the turn region; Ile, Ser, and Phe in the second helix, which is thought to make sequence-specific contacts. Sequence identities in the second helix of both regions of Mu A include: Glu followed by Tyr, and the penultimate Pro.

homologous region of Mu A (residues 389–408) with a predicted α-helix–turn–α-helix structural unit is also aligned in Fig. 5.

Mu Transposition. Knowing how proteins recognize specific base sequences within double-stranded DNA is essential to understanding gene expression. The CAP, cro, and cI crystallographic structures have helped illuminate the mechanism of protein–DNA recognition. The Mu transposase, unlike these proteins of known structure, performs functions more complex than simple binding to the DNA. Therefore, Mu A regions that show homology with this family of sequences and appear to satisfy requirements for adopting this bihelical conformation are of considerable interest.

The predicted bihelical motif in the Mu repressor protein (Fig. 5) has a conserved glutamine residue at position 1, as do most other phage repressor sequences (12). Since sequence homology between the repressor and Mu A is extensive in this bihelical region, the corresponding sequence in Mu A (residues 35–55) may also be involved in operator binding. Homologies between Mu A and repressor in this region differ from those between Mu A and D108 A. This is consistent with recognition by Mu and D108 A of the same sequence for transposition but probably different operator sequences. We suggest that the same region in these transposases binds the two DNA sequences (i.e., operator and ends) but makes some contacts with these two sets of sequences through different sets of amino acid residues.

Sequence homologies between the two regions of Mu A transposase aligned in Fig. 5 (residues 35–55 and 389–408) suggest that the second region contains a binding site for Mu DNA ends, as proposed for the first region. The two ends of Mu function nonequivalently in transposition—i.e., the efficiency of transposing two left or two right ends is at least two orders of magnitude lower than that of transposing one left and one right end (unpublished results). Perhaps the two ends are recognized by two different binding sites on the transposase, ensuring that only the left–right end combination can initiate transposition efficiently.

Sequence homologies between D108 A and Mu B (Fig. 5) in their predicted DNA-binding regions may imply that B also has a role to play in recognizing the DNA ends. Although it does not bind specifically to the ends, B does seem to bind DNA nonspecifically (15). Since Mu can transpose into nearly random locations in the *E. coli* genome (1), perhaps one function of B is to assist the Mu ends in binding random target sites on the DNA.

Conclusions. The Mu transposase and repressor proteins and the D108 transposase have significant sequence similarities and also a region in common that could serve a DNA-binding function and possibly make contacts with two sets of DNA sequences (the repressor–operator and the phage ends). The presence of a second potential DNA-binding site in the Mu transposase suggests that the Mu DNA ends may be bound at two sites on this protein. Finally, we note intriguing sequence similarities between the possible DNA-binding regions in D108 transposase and Mu transposition protein B.

We thank George Fey and his group, Steve Anderson, Jan Scal, and Marteen de Bruijn, for help with parts of the sequencing, Dan Bloch for protein searches, D. E. McRee for the program SECSTR, C. Nakayama for protein purification, and John Tainer for discussion. Kiyoshi Mizuuchi independently noticed homology between the published sequences of *A* and *c* (personal communication). J.L.M. has a studentship from the Medical Research Council of Canada. This work was supported by National Institutes of Health Grant GM 33247-01 to R.M.H. and grants from the National Institute of Cancer and Medical Research Council of Canada to G.C.

1. Bukhari, A. I. (1976) *Annu. Rev. Genet.* **10,** 389–412.
2. Toussaint, A. & Resibois, A. (1983) in *Mobile Genetic Elements*, ed. Shapiro, J. (Academic, New York), pp. 105–158.
3. Kahmann, R. & Kamp, D. (1979) *Nature (London)* **280,** 247–250.
4. Craigie, R., Mizuuchi, M. & Mizuuchi, K. (1984) *Cell* **39,** 387–394.
5. Giphart-Gassler, M., Reeve, J. & van de Putte, P. (1981) *J. Mol. Biol.* **145,** 165–191.
6. Craigie, R. & Mizuuchi, K. (1985) *J. Biol. Chem.* **260,** 1832–1835.
7. Goosen, N., van Heuvel, M., Moolenaar, G. F. & van de Putte, P. (1984) *Gene* **32,** 419–426.
8. Gill, G. S., Hull, R. C. & Curtiss, R., III (1981) *J. Virol.* **37,** 420–430.
9. Toussaint, A., Faelen, M., Desmet, L. & Allet, B. (1983) *Mol. Gen. Genet.* **190,** 70–79.
10. Priess, H., Kamp, D., Kahmann, R., Brauer, B. & Delius, H. (1982) *Mol. Gen. Genet.* **186,** 315–321.
11. Ohlendorf, D. H., Anderson, W. F. & Matthews, B. W. (1983) *J. Mol. Evol.* **19,** 109–114.
12. Pabo, C. O. & Sauer, R. T. (1984) *Annu. Rev. Biochem.* **53,** 293–321.
13. Chaconas, G., deBruijn, F. J., Casadaban, M. J., Lupski, J. R., Kwoh, T. J., Harshey, R. M., Dubow, M. S. & Bukhari, A. I. (1981) *Gene* **13,** 37–46.
14. Miller, J. L., Anderson, S. K., Fujita, D. J., Chaconas, G., Baldwin, D. L. & Harshey, R. M. (1984) *Nucleic Acids Res.* **12,** 8627–8638.
15. Chaconas, G., Gloor, G. & Miller, J. L. (1985) *J. Biol. Chem.* **260,** 2662–2669.
16. Harshey, R. M. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 2012–2016.
17. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
18. Bankier, A. T. & Barell, B. G. (1983) in *Techniques in Nucleic Acid Biochemistry*, ed. Flavell, R. A. (Elsevier/North-Holland, Limerick, Ireland), Vol. B5-08, pp. 1–34.
19. Staden, R. (1982) *Nucleic Acids Res.* **10,** 4731–4751.
20. Chou, P. Y. & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol.* **47,** 45–148.
21. Eisenberg, D., Weiss, R. M., Terwilliger, T. C. & Wilcox, W. (1982) *Faraday Symp. Chem. Soc.* **17,** 109–120.
22. Craig, N. L. & Nash, H. A. (1984) *Cell* **39,** 707–716.
23. Miller, H. I. & Friedman, D. I. (1980) *Cell* **20,** 711–719.
24. Groenen, M. A. M., Timmers, E. & van de Putte, P. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 2087–2091.
25. Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. (1981) *Nature (London)* **290,** 754–758.
26. Pabo, C. O. & Lewis, M. (1982) *Nature (London)* **298,** 443–447.
27. McKay, D. B., Weber, I. T. & Steitz, T. A. (1982) *J. Biol. Chem.* **257,** 9518–9524.
28. Sauer, R. T., Yocum, R. R., Doolittle, R. F., Lewis, M. & Pabo, C. O. (1982) *Nature (London)* **298,** 447–451.
29. Steitz, T. A., Ohlendorf, D. H., McKay, D. B., Anderson, W. F. & Matthews, B. W. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 3097–3100.
30. Dickerson, R. E. & Geis, I. (1983) *Hemoglobin: Structure, Function, Evolution and Pathology* (Benjamin/Cummings, Menlo Park, CA), pp. 68–70.
31. Weber, P. C., Howard, A., Xuong, N. H. & Salemme, F. R. (1981) *J. Mol. Biol.* **153,** 399–424.