Dictyostelium discoideum Nuclear Plasmid Ddp5 Is a Chimera Related to the Ddp1 and Ddp2 Plasmid Families

W. Kurt Rieben, Jr., Chad M. Gonzales, Shane T. Gonzales, Katharine J. Pilkington, Hidenori Kiyosawa,[1] Joanne E. Hughes and Dennis L. Welker

Molecular Biology Program, Biology Department, Utah State University, Logan, Utah 84322-5305

Manuscript received May 1, 1997
Accepted for publication November 24, 1997

ABSTRACT

The 14,955-bp *Dictyostelium discoideum* nuclear plasmid Ddp5 contains six transcribed open reading frames. One of these is related to the *rep* gene of the Ddp2 plasmid, and the other five are related to genes present on the Ddp1 plasmid. The absence of a homolog of the Ddp1 G1 gene, coupled with the presence of the Ddp2 *rep* gene homolog and of a 1.6-kb inverted repeat analogous to the inverted repeats on members of the Ddp2 plasmid family, suggests that Ddp5 uses Ddp2-like replication and copy number control mechanisms and that it should be assigned to the Ddp2 plasmid family. Ddp5 carries genes homologous to the D1/D3 and D2 genes of the Ddp1 plasmid as well as the Ddp1 G2/G3/D4, G5/D6, and G6/G4/D5 genes. The products of the Ddp5 G2-like, G5-like, and G6-like genes are likely to be transcription factors regulating the expression of themselves and of the other Ddp5 genes. The D1-like and D2-like genes may confer a selective advantage to plasmid-bearing cells, because they can be deleted from plasmid-based shuttle vectors with no apparent effect on vector maintenance. Updated sequence information for the Ddp1 G5/D6, D1/D3, and D2 genes as well as the Dmp1 and Dmp2 G5-like genes is presented. The locations of introns in the G5-like and D1-like genes of Ddp5 and in the homologous genes of the Ddp1, Dmp1, and Dmp2 plasmids were identified. These introns all have GU at the 5′ intron border and AG at the 3′ intron border, are short (59 to 71 nucleotides), and are AT-rich. A conserved HHCC domain was identified in the G5 proteins; this is a putative zinc binding domain and may be involved in protein-DNA interaction.

UNLIKE other eukaryotes, cellular slime molds in the genus Dictyostelium contain a diverse set of circular, high copy number, nuclear plasmids (Metz *et al.* 1983; Noegel *et al.* 1985; Hughes *et al.* 1988). The focus of our recent work has been to identify relationships between these plasmids (Yin and Welker 1992; Kiyosawa *et al.* 1993; Kiyosawa *et al.* 1994; Shammat *et al.*, 1998), to identify the functions of plasmid genes and sequence elements (Hughes *et al.* 1992; Hughes *et al.* 1994; Kiyosawa *et al.* 1995), and to develop shuttle vectors from these plasmids (Hughes *et al.* 1992; Hughes *et al.* 1994; Shammat *et al.*, 1998). Comparison of plasmid sequences allows conserved and nonconserved features to be identified and later targeted in function studies. At least four distinct plasmid families exist based on shared gene and structural features. These are the Ddp1, Ddp2, Dpp1, and Dpp3 families, which are named after representative plasmids. Gene function studies have focused on Ddp1 (Hughes *et al.* 1994; Kiyosawa *et al.* 1995), Ddp2 (Chang *et al.* 1990; Leiting *et al.* 1990; Slade *et al.* 1990; Hughes *et al.* 1992), and Ddp6 (Shammat *et al.* 1998). The genes on these plasmids, with the exception of the Ddp1 D1/D3 and D2 genes, have been shown to play critical roles in plasmid maintenance and copy number control, because their inactivation leads to rapid loss of plasmid-based shuttle vectors under nonselective growth conditions (Hughes *et al.* 1992; Hughes *et al.* 1994; Kiyosawa *et al.* 1995; Shammat *et al.* 1998).

The plasmid Ddp5 has previously been shown to be present at high copy number in *Dictyostelium discoideum* wild isolate WS2162 (Noegel *et al.* 1985), located in the cell nucleus based on its cofractionation with nuclei and its chromatin structure (Ashktorab and Welker 1988), and compatible with the Ddp1 and Ddp2 plasmids when the native plasmids are present in the same cell (Hughes and Welker 1989). In this article we present the complete Ddp5 sequence, identify its surprising relationship to both Ddp1 and Ddp2, and discuss the functions of the Ddp5 gene products. We also provide updated sequence information that alters the predicted protein products for three Ddp1 genes and one gene

Corresponding author: Dennis L. Welker, Biology Department, Utah State University, Logan, UT 84322-5305.
E-mail: dlw@cc.usu.edu

[1] *Present address:* Department of Cell Science, Institute of Biomedical Sciences, Fukushima Medical College, Hikari-gaoka 1, Fukushima-shi, Fukushima-ken, Japan 960-12.

from each of the plasmids Dmp1 and Dmp2. The implications of these findings for further analysis of plasmid gene functions are discussed.

MATERIALS AND METHODS

DNA, cDNA, and protein: Native Ddp5 plasmid DNA was isolated from the nuclei of cells of the wild isolate WS2162 using a modification of the technique of Birnboim and Doly (1979; Metz *et al.* 1983; Hughes *et al.* 1988; Hughes and Welker 1989). This DNA was cloned into the *Escherichia coli* vectors pGEM3Z, pGEM4Z, pGEM5Z, or pGEM7Z. More subclones were made either by exonuclease III digestion and religation or by cloning of smaller restriction fragments. The p94d5 vector carries all of the Ddp5 sequence, except the region between the *Xba*I site at position 4655 and the *Hin*dIII site at position 9710, cloned into the *Xba*I and *Hin*dIII sites of the pGEM3Z vector. DNA sequences were obtained by the dideoxynucleotide method through the Utah State University Biotechnology Center. The plasmid protein comparisons used the GCG Gap, Lineup and Pileup Programs with their default settings. DNA and protein genebanks were searched for matches to the plasmid proteins using the programs available at the National Center for Biotechnology Information website. cDNA fragments were obtained after RT-PCR using pairs of oligonucleotide primers specific for the plasmid genes, and then they were cloned and sequenced, and the sequences compared to that of the corresponding genomic plasmid DNA. The accession numbers of the plasmids are the following: Ddp2, M55298, and X51478; Ddp5, AF000580; Ddp1, U00691, and U00796; Dmp1, U00175; Dmp2, U00176.

Total and poly A+ RNA: For preparation of Ddp5 transcripts, axenically grown *D. discoideum* cells of the Ddp5 transformant HUD896 or bacterially grown cells of the wild isolate WS2162 were harvested in vegetative growth at aggregation or at culmination during the asexual life cycle and fruiting body formation. RNA preparations were also obtained from bacterially grown cells of the *D. discoideum* wild isolate NC4 (Ddp1) and of the *Dictyostelium mucoroides* wild isolate DMUC2 (Dmp1 and Dmp2). Total RNA was isolated either by following the protocol of Franke and coworkers (Franke *et al.* 1987) or by using Tri Reagent (Molecular Research Center Inc., Cincinnati, OH) (Kiyosawa *et al.* 1994, 1995). To isolate poly A+ RNA either a Poly A Quik mRNA Purification Kit (Stratagene, La Jolla, CA) or a PolyATtract mRNA Isolation System (Promega, Madison, WI) was used. Glyoxylated RNA was separated on 1% agarose gels, blotted to nylon membrane, and probed with ^{32}P-labeled DNA fragments.

RESULTS

General features of Ddp5: A circular map of the 14,955-bp Ddp5 sequence showing the relative locations of selected restriction sites, genes, and repeat elements is presented in Figure 1. Ddp5 carries six long open reading frames (ORFs). The 25.4% G + C content of the plasmid is similar to that for Dictyostelium chromosomal DNA. The six transcribed ORFs, as found in the mRNAs, have a combined G + C content of 27%, and the intergenic and intronic regions have a combined

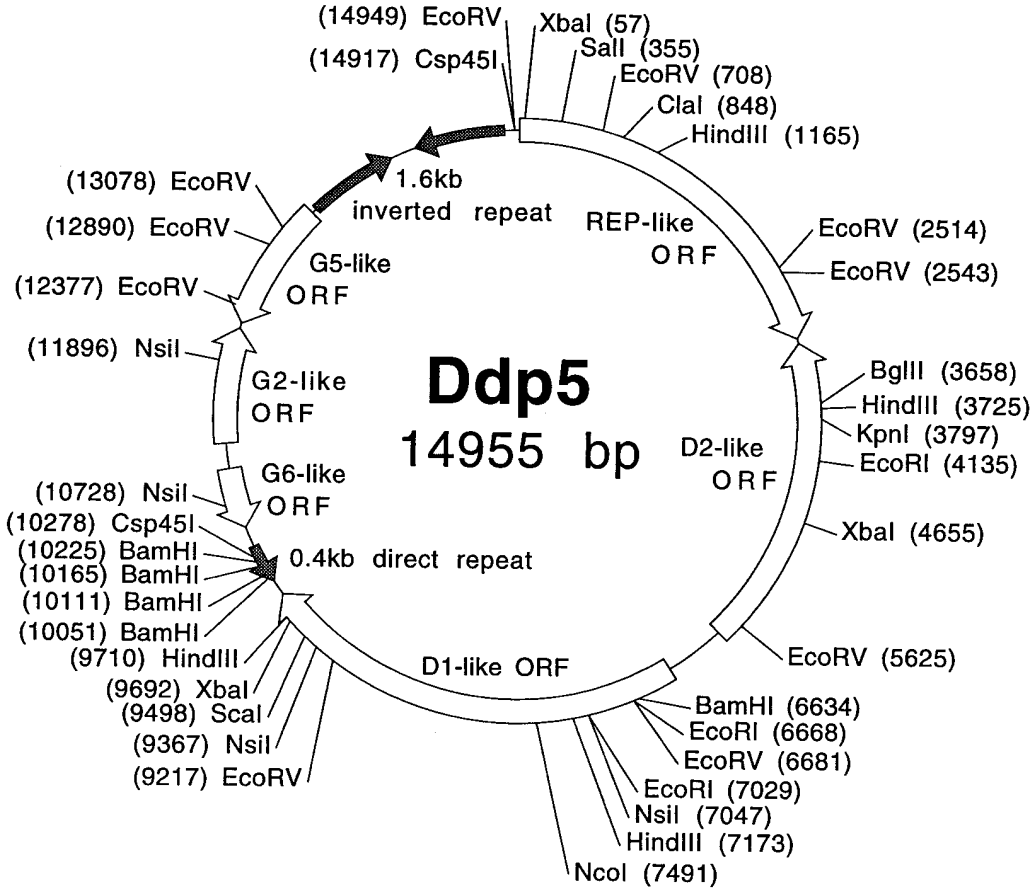

Figure 1.—Map of the Ddp5 plasmid showing the positions of selected restriction sites, ORFs, and repeats.

TABLE 1

Transcripts of *D. discoideum* plasmid Ddp5

Gene	Size (kb)	ORF (bp) length	Vegetative growth	Aggregation stage	Culmination stage
D1-like	4.1	3822	−	−	++
D2-like	2.8	2604	−	+	++
G2-like	1.1	933	+?	++	++
G5-like	1.2	1020	+	++	++
G6-like	0.6	513	−	+	++
rep-like	3.4	3114	++	++	−

A plus sign means the transcript was detected, two plus signs indicate the stage at which maximal transcription was detected, and a negative sign means the transcript was not detected. The presence of the G2-like gene transcript in vegetative cells is uncertain.

G + C content of 19%. The six coding regions (including the introns) comprise 81.3% of the sequence, with the majority of the noncoding sequences present in the repeat elements.

Relationship of Ddp5 with the Ddp2 plasmid family: The presence of a long inverted repeat is a characteristic feature of the Ddp2 plasmid family. For example, the inverted repeat of Ddp2 spans a 1-kb region with its 0.5-kb repeat elements separated by a short sequence of DNA. Within the inverted repeat region are additional smaller inverted and direct repeat structures. The Ddp2 repeat lies upstream of the promoter for the plasmid's *rep* gene, which is the only gene on Ddp2. Three other members of the Ddp2 plasmid family (pDG1, Dgp1, and Ddp6) are organized in an identical way. Ddp5's inverted repeat spans 1.6 kb and lies upstream of a gene with significant similarity to the *rep* genes of the plasmids in the Ddp2 plasmid family. The elements of the Ddp5 inverted repeat are 0.7 kb in length (689 and 705 bp) and are separated by 225 bp. The Ddp5 repeat units differ at a few positions because of base substitution and deletion mutations. The central 225 bp contain features of both inverted and direct repeats. Within it the palindromic repeat element CAGTCAGACTG occurs three times, and the closely related element CAGTCCGACTG occurs twice. Additional small inverted and direct repeats are also present in the long inverted repeat.

The Ddp5 *rep* gene is transcribed during vegetative growth and during fruiting-body formation at aggregation but not at culmination (Table 1). The predicted protein product of this gene retains significant similarity in sequence and presumably in structure and function to those of the *rep* genes of the other members of the Ddp2 plasmid family. For example, there is 52% similarity and 31% identity with the Ddp2 Rep protein. The Ddp5 protein retains all of the features noted previously in comparisons of the Ddp2 family's Rep proteins except a polythreonine motif located approximately 250 residues into the peptide sequence (Figure 2). In the Ddp5 Rep protein this region is expanded, and there is a polyaspartic acid motif nearby. This is one of the two most divergent regions of the Ddp5 Rep protein. It is known that this region of the Ddp2 Rep protein can be removed with no apparent effect on protein function (Leiting *et al.* 1990). The most striking feature of the peptide sequences of the Rep proteins is their acidic carboxy termini. This feature is retained but expanded in the Ddp5 Rep protein by about 50 residues. The differences in these two regions account for the larger size of the Ddp5 Rep protein (1038 amino acids) compared to the other Rep proteins (887 to 932 amino acids). The Rep proteins of the Ddp2 and Ddp6 plasmids have been shown to be critical for plasmid maintenance and copy number control (Leiting *et al.* 1990; Hughes *et al.* 1992; Shammat *et al.* 1998); it is highly likely that the Ddp5 Rep protein serves similar functions in the maintenance of Ddp5 (Table 2).

Relationship of Ddp5 with the Ddp1 plasmid family: Surprisingly, the remaining five Ddp5 genes are similar to five of the six genes present on the Ddp1 plasmid. The Ddp1 gene without a match is the G1 gene. The G1 gene is one of only two genes present on the Dmp1 and Dmp2 plasmids, the two other members of the Ddp1 plasmid family (Kiyosawa *et al.* 1994). Inactivation of the Ddp1 gene leads to plasmid multimerization, lowered copy number, and rapid loss of vector DNA (Hughes *et al.* 1992; Hughes *et al.* 1994). These features suggest that the G1 gene plays a key role in the replication control and maintenance of the plasmids in the Ddp1 family. Its absence from Ddp5, coupled with the presence of a Ddp2-like inverted repeat and *rep* gene, suggests that Ddp5 is utilizing Ddp2-like replication control and maintenance mechanisms.

Proceeding clockwise from the *rep* gene on the Ddp5 map, the next gene is similar to the D2 gene of Ddp1. The Ddp5 D2-like gene is transcribed during aggregation and culmination (Table 1). Comparison of the predicted protein products of the two genes shows that they are organized similarly, with the Ddp5 protein being larger—868 compared to 763 amino acids in the Ddp1 protein (Figure 3). The residues in the proteins are 51% similar and 26% identical. The Ddp1 data used

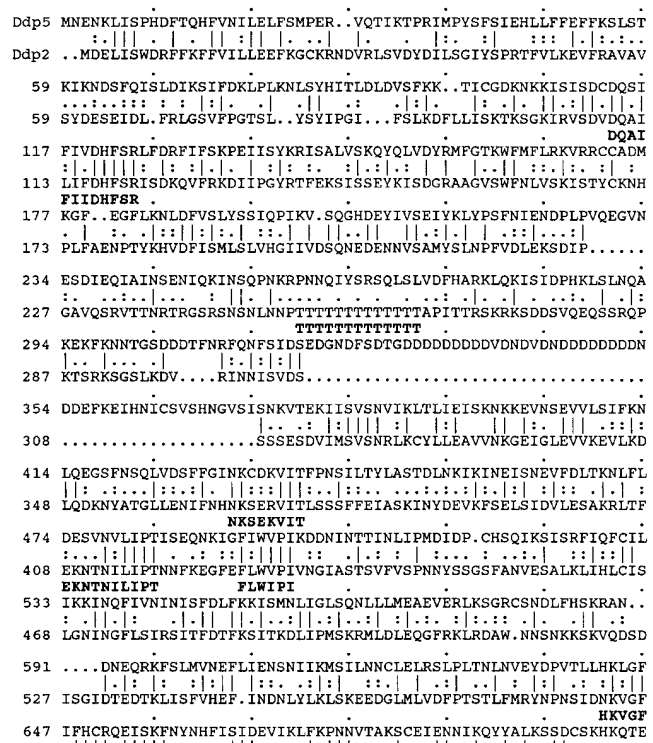

Figure 2.—Alignment of the Ddp5 and Ddp2 Rep proteins. Conserved features found in all five Rep proteins are indicated by the consensus sequences in bold under the aligned sequences. Because the consensus sequences are also based on the sequences of the Ddp6, Dgp1, and pDG1 Rep proteins, they differ in some positions from the sequences shown for the Ddp5 and Ddp2 Rep proteins.

in this comparison differs from that previously published (Farrar *et al.* 1994) in having an additional nucleotide pair inserted near the 3′ end of the gene. This increases the size of the protein and alters its carboxy terminus. The function of the Ddp1 D2 protein and its Ddp5 homolog are not known. Deletion of the Ddp1 D2 gene has little or no effect on vector maintenance (Hughes *et al.* 1994). Similarly, the inactivation of the Ddp5 D2-like gene had no apparent effect on vector maintenance. The Ddp5-based vector p94d5, which has the region of Ddp5 from the *Xba*I site at position 4655 to the *Hin*dIII site at position 9710 deleted, replicated autonomously at high copy number and was stably maintained for 400 generations of growth under nonselective conditions. The p94d5 vector lacks a functional D2-like gene, as well as a functional D1-like gene. Because the D2-like gene product appears to not be involved in plasmid maintenance, it is possible that it confers a selective advantage to plasmid-bearing cells (Table 2).

Clockwise from the D2-like gene is a gene with similarity to the D1/D3 gene of Ddp1. The Ddp5 gene is transcribed at culmination (Table 1). Both the Ddp5 and the Ddp1 genes are split into two exons with a short intron near the 5′ end of the gene. The Ddp5 intron splits codon 26 (isoleucine) and the Ddp1 intron splits codon 33 (lysine). The intron borders follow the GU-AG rule. The intron in the Ddp5 gene is 70 nucleotides in length and that in the Ddp1 gene contains 62 nucleotides. Because the presence of the Ddp1 intron has not been previously reported, the predicted Ddp1 D1 gene product used in this work has an additional 54 amino acids at its amino terminus from that reported elsewhere (Farrar *et al.* 1994). There are 1274 amino acids in the Ddp5 protein and 1474 in the Ddp1 protein (Figure 4). The amino acid residues are 48% similar and 25% identical. The Ddp1 protein has been proposed to be a transcription factor based on the presence of a putative leucine zipper domain two-thirds of the way through the protein and of a putative zinc finger domain near its carboxy terminus (Farrar *et al.* 1994). The Ddp5 protein appears to lack a leucine zipper domain. At the carboxy end of the Ddp5 protein there is a cysteine-rich region analogous to the putative zinc finger domain of the Ddp1 protein. There are three possible zinc fingers in each protein. Deletion of the Ddp1 D1/D3 gene had no apparent effect on maintenance of Ddp1-based transformation vectors (Hughes *et al.* 1994). Similarly, deletion of the Ddp5 D1-like gene had no apparent effect on maintenance, autonomous replication, or copy number of the p94d5 vector. The gene most likely to be regulated by the presence of a D1/D3 transcription factor is the plasmid's D2 gene. Alternatively, the D1/D3 gene product may work with the D2 gene product to confer a selective advantage to plasmid-bearing cells (Table 2).

After the D1-like gene are five copies of a degenerate 60-bp direct repeat, with portions of two additional copies. Prominent features of the repeat are the presence of a *Bam*HI or a *Csp*45I restriction site and of the sequence TGAAACTTTTAAAA. There is a 306-nucleotide ORF that overlaps the repeat. We have no evidence for transcription through this region, and the repetitive nature of the region suggests it is not expressed.

The next gene on Ddp5 has similarity to the G6/G4/D5 gene of Ddp1. The Ddp5 gene is transcribed during aggregation and culmination (Table 1). The Ddp5 protein corresponds best to the amino terminal end of the Ddp1 G6 protein (Figure 5). The Ddp1 protein has 254 amino acids, while the Ddp5 protein has only 171 amino

TABLE 2

Functions of plasmid genes

Ddp5 gene	Homologous gene	Comments on functions of the homologous Ddp1 and Ddp2 genes
D1-like	Ddp1 D1/D3	No known function, not involved in plasmid maintenance, may provide selective benefit to cells (Hughes *et al.* 1994), may encode a transcription factor regulating the D2 gene.
D2-like	Ddp1 D2	No known function, not involved in plasmid maintenance, may provide selective benefit to cells (Hughes *et al.* 1994).
G2-like	Ddp1 G2/G3/D4	Required for maintenance of Ddp1 (Hughes *et al.* 1994), may encode a transcription factor.
G5-like	Ddp1 G5/D6	Required for maintenance of Ddp1, encodes a transcription factor that negatively regulates G6 transcript levels, may play other roles in transcription and replication control (Hughes *et al.* 1994; Kiyosawa *et al.* 1995).
G6-like	Ddp1 G6/G4/D5	Required for maintenance of Ddp1, encodes a transcription factor that regulates G2/G3/D4 gene transcription, may play other roles in transcription and replication control (Hughes *et al.* 1994).
rep-like	Ddp2 *rep*	Required for long-term maintenance of Ddp2, controls copy number, and likely regulates replication when copy numbers fall too low (Chang *et al.* 1990; Leiting *et al.* 1990; Hughes *et al.* 1992).

acids. The predicted proteins are 49% similar and 22% identical. If the Ddp1 D5 protein, which comes from the same ORF using an alternative start codon, is used in the comparison, the peptides are only 45% similar and 21% identical. The Ddp1 gene products appear to be transcription factors because disruption of the gene alters expression of other Ddp1 genes (Hughes *et al.* 1994). The Ddp5 gene product is likely to play a similar role (Table 2).

The next gene on the Ddp5 plasmid has similarity to the Ddp1 G2/G3/D4 gene. The Ddp5 gene is transcribed during aggregation and culmination and may be transcribed during vegetative growth (Table 1). The predicted protein products of the Ddp5 and Ddp1 genes are 51% similar and 25% identical. The Ddp5 protein has 311 amino acids compared to 401 in the Ddp1 protein (Figure 6). The Ddp5 protein corresponds best to the amino terminus of the Ddp1 protein. Disruption of the Ddp1 G2/G3/D4 gene led to multimerization of Ddp1-based plasmid vectors and decreased plasmid maintenance (Hughes *et al.* 1994). It is probable from the absence of G2-like genes from the Dmp1 and Dmp2 plasmids that the G1 gene product is the key protein in controlling multimer formation, with the G2 Ddp1 gene products serving to regulate G1 transcription. It is likely that the Ddp5 G2-like gene also encodes a transcription factor (Table 2).

The last gene on Ddp5 is similar to the G5/D6 gene of Ddp1 and the G5-like genes of the Dmp1 and Dmp2 plasmids. The Ddp5 gene appears to be transcribed at a low level during vegetative growth and at a higher level during aggregation and culmination (Table 1). The predicted protein products of the Ddp5 and Ddp1 genes are 42% similar and 24% identical. The Ddp5 protein is 61% similar and 31% identical with the Dmp1 protein. The proteins are organized similarly and are of similar length (Figure 7). There are 340 amino acids in the Ddp5 protein, 325 in the Ddp1 protein, 264 in the Dmp1 protein, and 275 in the Dmp2 protein. All four proteins contain an acidic motif rich in glutamic acid and aspartic acid residues near the middle of the protein. This may serve as an activation domain for interaction with and control of cellular transcription and replication factors (Table 2). The Ddp1, Dmp1, and Dmp2 proteins used in this comparison differ from those originally reported (Farrar *et al.* 1994; Kiyosawa *et al.* 1994). All four genes are interrupted by single introns. The Ddp5 intron splits codon 95 (glycine), the Ddp1 intron splits codon 85 (glycine), and the Dmp1 and Dmp2 introns split codon 92 (aspartic acid). The intron borders all follow the GU-AG rule. The Ddp5 intron contains 71 nucleotides, the Ddp1 intron contains 59 nucleotides, the Dmp1 intron contains 61 nucleotides, and the Dmp2 intron contains 60 nucleotides. In addition, sequencing of Ddp1 G5/D6 cDNA and genomic DNA clones has revealed the presence of a 54-bp *Pst*I fragment that was not reported earlier (Gurniak *et al.* 1990; Farrar *et al.* 1994; Kiyosawa *et al.* 1994). The Ddp1 G5/D6 gene product is known to negatively regulate the G6 transcript (Kiyosawa *et al.* 1995), and it appears to be a positive regulator of other Ddp1 transcripts (Kiyosawa *et al.* 1995). The Ddp5 G5-like gene product is likely to be a transcription factor as well (Table 2).

Comparison of the G5 proteins identified a new feature that may be of considerable importance to protein function. This is a putative zinc finger motif that is among the most highly conserved features of the pro-

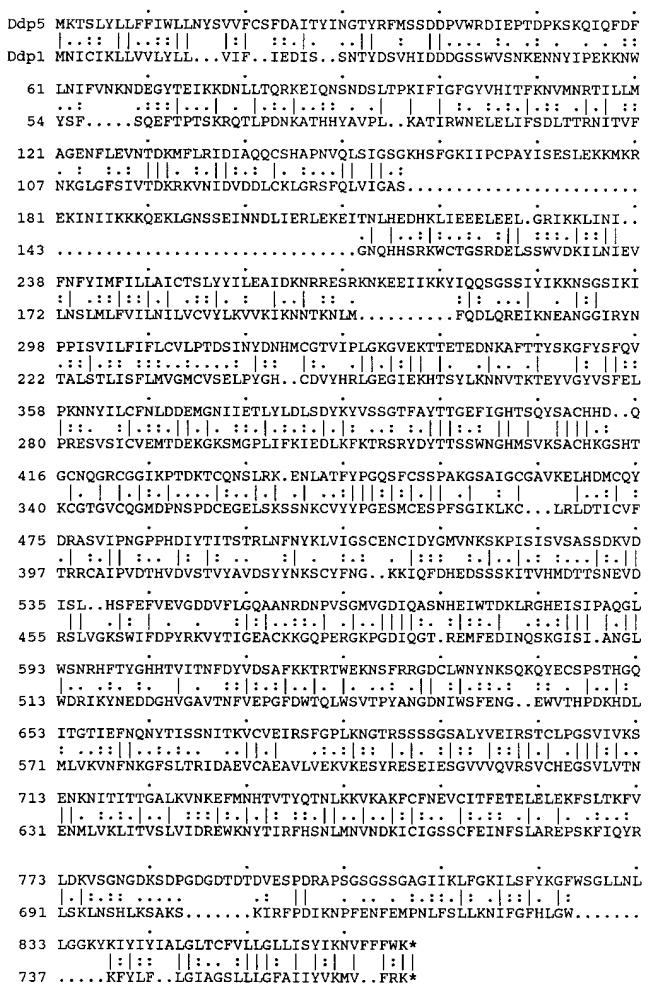

Figure 3.—Alignment of the Ddp5 and Ddp1 D2 proteins. The Ddp1 protein sequence is based on an updated Ddp1 sequence and differs at its carboxy terminus from that given elsewhere (Farrar *et al.*, 1994).

teins (Figure 7). The putative zinc binding domain has structural similarity to the HHCC domains of retroviral integrase proteins. These are known to bind zinc and are thought to be involved in DNA binding (Burke *et al.* 1992; Bushman *et al.* 1993; Vincent *et al.* 1993). In the integrase proteins the spacing between the central H and C residues is 23 to 32 amino acids, whereas the spacing in the G5 proteins is 28 amino acids. The C residues are separated in the integrase and the G5 proteins by 2 residues, and the H residues are separated by 3 to 7 amino acids in the integrase and 5 to 19 amino acids in the G5 proteins. The amino acids in the central H to C region are similar in the G5 proteins (Figure 7) but different from the amino acids found in the corresponding regions of the integrase proteins. The integrase proteins work as part of a multimeric complex, and it is likely that the G5 proteins do also, because a single zinc finger by itself is likely to be insufficient for DNA binding.

DISCUSSION

The Ddp5 plasmid has a complex genetic organization and is related to both the Ddp1 and Ddp2 plasmids. It appears to utilize a Ddp2-like maintenance mechanism, because it carries a gene related to the Ddp2 *rep* gene and Ddp2-like inverted repeat but lacks a Ddp1-like G1 gene. We therefore place Ddp5 in the Ddp2 plasmid family.

Ddp5 carries five genes related to the Ddp1 G2/G3/D4, G6/G4/D5, G5/D6, D1/D3, and D2 genes. It is intriguing to find these genes retained on Ddp5 when only G1-like and G5-like genes are present on the Dmp1 and Dmp2 plasmids of the Ddp1 plasmid family (Kiyosawa *et al.* 1994). Of the five genes, perhaps the most interesting are the D1-like and D2-like genes. As with the Ddp1 D1/D3 and D2 genes, the Ddp5 D1-like and D2-like genes are not required for plasmid maintenance. Deletion of these genes does not prevent long-term, extrachromosomal replication and maintenance of Ddp5-based vectors. The functions of the D1 and D2 genes remain to be determined; it is possible that this pair of genes provides a beneficial phenotype to plasmid-bearing cells. The Ddp5 G5-like, G6-like, and G2-like gene products are all likely to be transcription factors. Overall, the Ddp5 gene products have 22 to 26% identity and 42 to 51% similarity with those of Ddp1. This level of conservation is sufficient to allow the proteins to attain similar structures with related, but plasmid-specific, tertiary domains for protein-DNA and protein-protein interactions (Service 1997). Comparison of the Ddp5 proteins with those of the other plasmids, in particular Ddp1, reveals numerous conserved as well as nonconserved features. Because the Ddp5 peptide sequences are often only the second allele of these genes that has been studied, these comparisons provide new insight into the features likely to be important for protein function. This information, along with the revised sequences provided for the Ddp1 D1, D2, and G5/D6 gene products, will influence the direction of further analysis of the Ddp1 gene functions.

A significant new structural feature was identified in the G5 protein. This protein from Ddp1 and the related proteins from Ddp5, Dmp1, and Dmp2 contain an HHCC motif that may be involved in DNA binding. This motif is similar in size and organization to a zinc finger motif found in retroviral integrase proteins, which is known to bind zinc and is thought to be involved in binding to the retroviral LTRs (Burke *et al.* 1992; Bushman *et al.* 1993; Vincent *et al.* 1993).

The Ddp5 Rep protein is 29 to 34% identical and 52 to 55% similar to those of the other Ddp2 family members. Although it is the most divergent of the Rep proteins, it carries most of the previously identified conserved motifs; only the polythreonine box is not present. The Ddp2 and Ddp6 Rep proteins are important for plasmid maintenance and copy number control

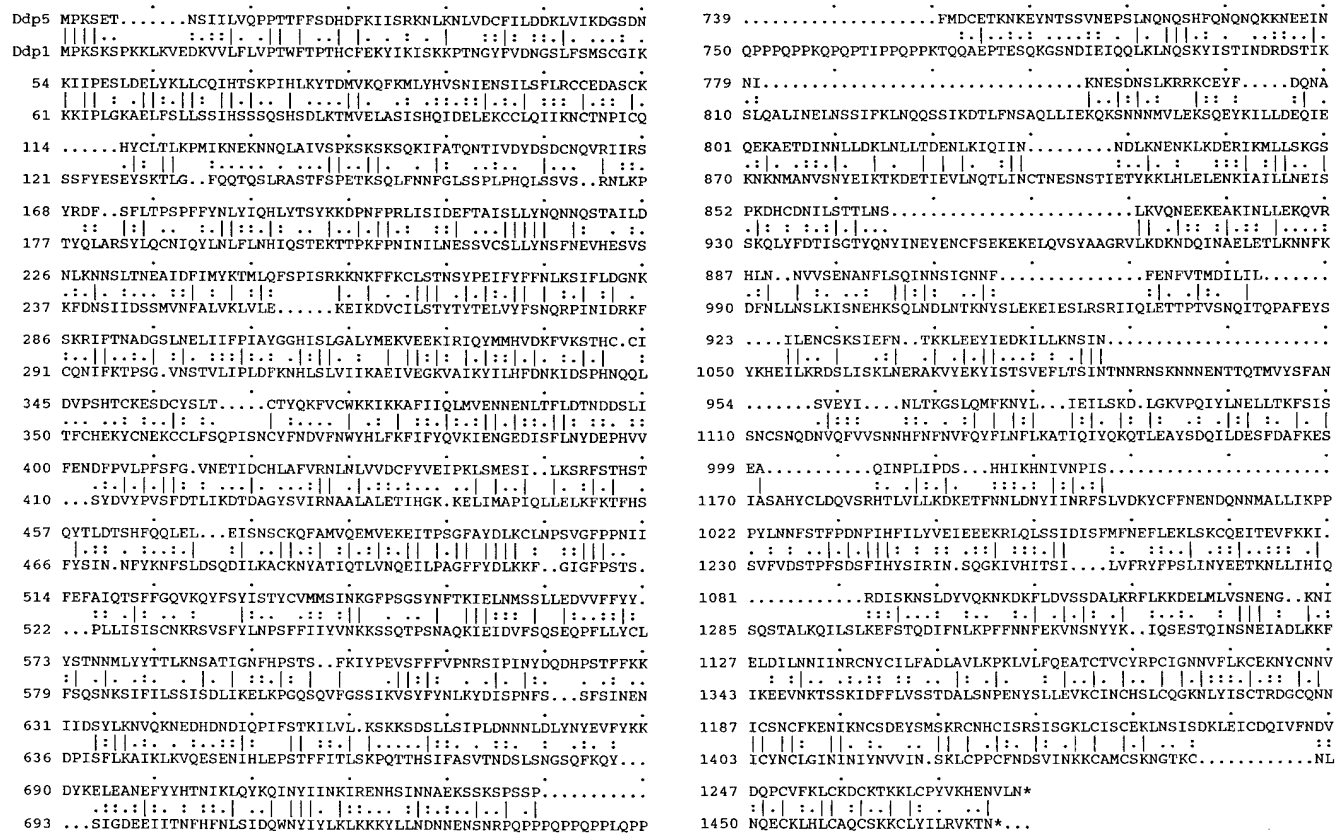

Figure 4.—Alignment of the Ddp5 and Ddp1 D1 proteins. The intron in the Ddp5 gene splits codon 26 (isoleucine) and the intron in the Ddp1 gene splits codon 33 (lysine). The last 110 amino acids at the carboxy termini of both proteins are rich in cysteine residues; these may be involved in zinc finger domains, although the spacing between some cysteine residues is different in the two proteins. The putative leucine zipper in the Ddp1 D1 protein between positions 1010 and 1031 appears not to be conserved in the Ddp5 D1 protein. The Ddp1 D1 protein differs from that given elsewhere (Farrar *et al.* 1994) by an addititonal 54 amino acids at its amino terminus.

(Hughes *et al.* 1992; Shammat *et al.* 1998). The Ddp2 Rep protein is also thought to be involved in negative transcriptional regulation of its own gene and plasmid replication (Chang *et al.* 1990; Leiting *et al.* 1990; Slade *et al.* 1990; Hughes *et al.* 1992). The Ddp5 Rep protein is expected to retain similar functions. Our current working hypothesis is that the Rep proteins contain at least three functional domains: a DNA binding domain, a multimerization domain for interactions with other Rep protein subunits, and an activation domain for interactions with cellular transcription and replication factors. The activation domain in the Rep proteins is likely to be the acidic carboxy termini that are present in all Rep proteins including that of Ddp5. In other work, we have shown that the Ddp5, Ddp2, and Ddp6 Rep proteins appear to form plasmid-specific multimers

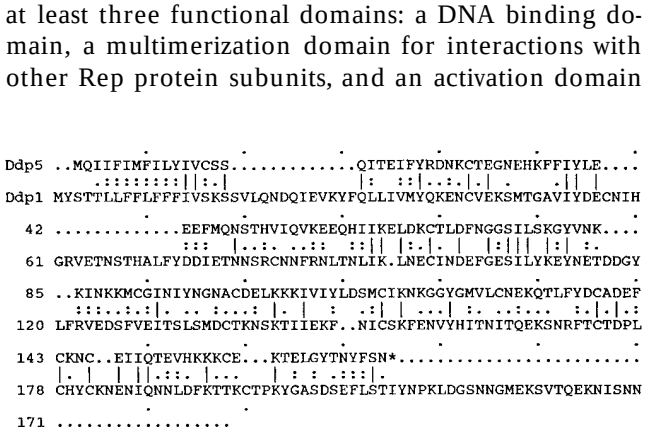

Figure 5.—Alignment of the Ddp5 and Ddp1 G6 proteins.

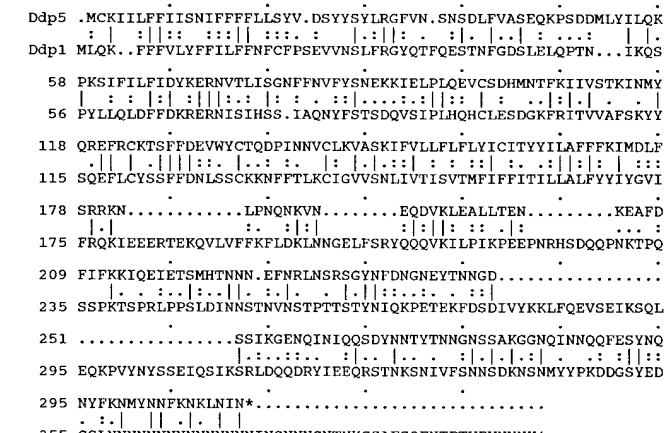

Figure 6.—Alignment of the Ddp5 and Ddp1 G2 proteins.

```
Ddp5        MESNYNEFYK NNLDIDHS.. QEIIRYLKNG NRVTVSNKHF FKDIFRSKFG
Dmp1        MDQETHNINF NNISQQLV.. DELIDVWSNK DKKKL..KRF LKPIYGCIY.
Ddp1        MIVVEIVFWT RIVKSKYSVD DEIFKYLVNR AKVVYNNSDY .KTLSQTDLE
CONSENSUS   M......f.. nn.....s.. dEii.Yl.N. .kv...nk.f .K.i......

Ddp5    49  DRTGRSKAY. .KSFVNSSFK RYLKIINKTY SLNGLDRNGD DILFNKIKGK
Dmp1    46  DRSTREYTLS RKSFINSTIK RFITNFFDNY IVKDVD..GV DVLYNKVRDL
Ddp1    50  KWAIQN.... ........TH RTITSIKNRF SVKKIGDEEK LFRISKNGEL
CONSENSUS   dr..r..... .ksf.ns..k R.it.i...y svk..d..g. d.l.nK...l

Ddp5    97  KVFIGPIPKR YHYN..EKDD NTI....... EHFNVDQMVK KIRDEGFYIC
Dmp1    94  KVYIGPISRK YHFVSFKRHP NTTITTPEPH SHISIGEMVK NIKREGVYVT
Ddp1    88  IVLNELEFDN FHIKEGK... .......... .HLRKSKMFN HIKDSGYYAT
CONSENSUS   kV.igpi... yH....k... nt........ .H.....Mvk .IkdeGyY.t
                        ----------------------------------------
Ddp5   138  KIDVENTFID CVDCQTKTYV VSK...KSTN LRKKTPDNEI HTPTLSPEQL
Dmp1   144  RTDIYNAFQG CIMC...... IAR...RRKD YVRKGDD... ..........
Ddp1   124  NEEIEIFLES CTLCKEITAQ TKRNSYKKRN IINKLPEEEE EEEEEEEEEE
CONSENSUS   ..dien.f.. C..C...t.. ..r...k..n ...K.pd.e. .......e..
            ---------------
Ddp5   185  TPLLTPQKTP PLSATLGEEI SEEETSEEET SEEETLEEET SPSISISNSP
Dmp1   172  .......KVT GLNMTQEGEM EEDDDDDDDD EEDE..EEEV PPP...RNSS
Ddp1   174  EEEEQEEEVE KPTISEEEEE ETPAVSEEEK EEEEEEEEET PAVSEEEKEE
CONSENSUS   .......kv. .l..t.eeE. ee...seee. eEeE..EEEt pp.....ns.

Ddp5   235  EVVIPLLPTP TPTPTPLISS PQSKVQQDQQ TSEGCLKFIP KLRYRNEIKA
Dmp1   210  RVLKSIPPKP ........SK PAGKKPAPSK PAATPTKLL. ..........
Ddp1   224  EEQEEDKEKD KEKKIEEDTE TGKKKDEVMN ERQDGIEISE K....TNDKP
CONSENSUS   ev.....pkp ........s. p..Kk..... ......k... k.......k.

Ddp5   285  SKLNKENTLE PTQKIIPVKK FKQSYWAKNP KLMSTTEIKV FQGCLSEKRS
Dmp1   241  .......... .......... ......GRNP KGISTSDISS MSKTMGNRRK
Ddp1   220  TEKTKVKKNV SRVKDINLHK KGNPITPKKI KNFSQQQLRL ASGVSNEKRA
CONSENSUS   ....k..... ...k.i...k .......knp K..St..i.. .sg...ekR.

Ddp5   335  RVQIKK*
Dmp1   265  *......
Ddp1   320  IITLKK*
CONSENSUS   ....kk*
```

Figure 7.—Alignment of the Ddp5, Dmp1, and Ddp1 G5 proteins. To avoid biasing the comparison, the Dmp2 G5-like protein was not included in this figure; it is 79% identical and 88% similar to the Dmp1 protein. The intron in the Ddp5 gene splits codon 95 (glycine), the introns in the Dmp1 and Dmp2 genes split codon 92 (aspartic acid), and the intron in the Ddp1 gene splits codon 85 (glycine). The G5 proteins contain a motif rich in glutamic and aspartic acid residues near the center of the proteins. The putative HHCC zinc finger motif (dashed line) is located between positions 108 and 151 in the Ddp5 protein, between positions 105 (or 112) and 157 in the Dmp1 protein, and between positions 99 and 137 in the Ddp1 protein. Retroviral integrase proteins have similar HHCC motifs with the spacing $HX_{3-7}HX_{23-32}CX_2C$ that are known to bind zinc and are thought to be involved in the interaction of the integrase with the viral LTRs (Burke *et al.* 1992; Vincent *et al.* 1992; Bushman *et al.* 1993).

in the yeast two-hybrid system and that the Ddp2 and Ddp6 Rep proteins have DNA binding capability (I. M. Shammat and D. L. Welker, unpublished results).

This is the first report of introns in Dictyostelium plasmid genes; however, the presence of the intron in the Ddp1 G5/D6 gene was first observed by Drs. Gurniak and Noegel (personal communication). Introns are present in the Ddp1 D1/D3 and G5/D6 genes, the Dmp1 and Dmp2 G5-like genes, and the Ddp5 D1-like and G5-like genes. The introns in related genes are located at similar positions near the 5′ ends of the genes. The introns mapped in this work follow the standard GU-AG rule for the 5′ and 3′ borders of the intron, are short, and are rich in AT base pairs. These features are typical of Dictyostelium introns.

Frequently asked questions concerning the Dictyostelium plasmids include: Where did they come from? What advantage, if any, do they provide to host cells? How do these plasmids function? Almost all plasmid genes that have been analyzed are involved in maintaining the plasmids in cells (Chang *et al.* 1990; Leiting *et al.* 1990; Hughes *et al.* 1992; Hughes *et al.* 1994; Kiyosawa *et al.* 1995; Shammat *et al.* 1998). This appears to be the case for most of the Ddp5 genes based on their similarity to genes known to possess this function. The only Dictyostelium plasmid genes that do not appear to be related to this function are the Ddp5 and Ddp1 D1 and D2 genes. These genes may provide selective advantages to plasmid-bearing cells. However, the phentotype that is affected by these genes is unknown. The GC contents of all sequenced Dictyostelium plasmids match that found for the Dictyostelium genome, that is, highly biased toward AT base pairs; this is consistent with an origin in Dictyostelium cells. Dictyostelium cells inhabit forest soils and phagocytize other microorganisms, many of which are likely to produce antimicrobial agents. Plasmid genes may provide defenses against such chemicals. The plasmids may have arisen from viruses (or still be proviruses) that provided protection from subsequent infection by similar viruses. Proviruses may evolve into "selfish DNAs" by loss of the ability to form viral particles while attaining efficient copy number control and maintenance mechanisms to ensure transmission to daughter cells. The plasmids present in fungal mitochondria provide clear examples of plasmid DNAs where viral features have been conserved (Griffiths 1995). The Dictyostelium life cycle has several points where differing cell fates determine transmission of genetic information to succeeding generations. In asexual fruiting-body formation, only spore cells produce progeny. The stalk cells that comprise about 20% of the initial population die. In the sexual cycle, which involves macrocyst formation, only the zygote produces progeny; all the other cells are phagocytosed and used as a food source by the zygote. Plasmid genes influencing cell fate during these processes would obviously improve the chances of plasmid maintenance in the population and lead to the preferential survival of plasmid-bearing cells. It is known that in *Absidia glauca* that a small circular extrachromosomal DNA carries the gene for a mating-type-specific cell surface protein (Hanfler *et al.* 1992). Dictyostelium are also capable of parasexual genetic exchange. The frequency of parasexual cell fusion may be increased by the presence of plasmids, thus leading to the spread of plasmids through cellular slime mold populations by this mechanism. Different Dictyostelium plasmids may influence one or more of these characteristics.

We thank our colleagues Drs. C. B. Gurniak and A. A. Noegel of the Max-Planck-Institut for Biochemistry for their personal communication concerning the presence of an intron in the Ddp1 G5 gene. We also thank Drs. G. Knapp and G. J. Podgorski for their advice. C.G., S.G., and K.P. were supported by undergraduate research fellowships under the auspices of a grant from the Howard Hughes Medical Institute to the Utah State University (USU) Biology Department. S.G. and K.P. were also supported by USU Undergraduate Research and Creative Opportunities fellowships. This work was supported by

funds from Utah State University and National Science Foundation grant DMB8915591.

LITERATURE CITED

Ashktorab, H., and D. L. Welker, 1988 Establishment of the nuclear location of *Dictyostelium discoideum* plasmids. Gene **65:** 41–49.

Birnboim, H. C., and J. Doly, 1979 A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7:** 1513–1523.

Burke, C. J., G. Sanyal, M. W. Bruner, J. A. Ryan, R. L. Lafemina *et al.*, 1992 Structural implications of spectroscopic characterization of a putative zinc finger peptide from HIV-1 integrase. J. Biol. Chem. **267:** 9639–9644.

Bushman, F. D., A. Engelman, I. Palmer, P. Wingfield and R. Craigie, 1993 Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. Proc. Natl. Acad. Sci. USA **90:** 3428–3432.

Chang, A. C. M., M. B. Slade and K. L. Williams, 1990 Identification of the origin of replication of the eukaryote *Dictyostelium discoideum* nuclear plasmid Ddp2. Plasmid **24:** 208–217.

Farrar, N. A., H. Kiyosawa, J. E. Hughes, D. L. Welker and K. L. Williams, 1994 Nucleotide sequence of Ddp1, a high copy number nuclear plasmid of *Dictyostelium discoideum*. Plasmid **31:** 184–195.

Franke, J., G. J. Podgorski and R. H. Kessin, 1987 The expression of two transcripts of the phosphodiesterase gene during the development of *Dictyostelium discoideum*. Dev. Biol. **124:** 504–511.

Griffiths, A. J. F., 1995 Natural plasmids of filamentous fungi. Microbiol. Rev. **59:** 673–685.

Gurniak, C. B., A. G. Bang and A. A. Noegel, 1990 Transcript and sequence analysis of a 5.1 kb contiguous fragment of *Dictyostelium discoideum* plasmid Ddp1 that contains the origin of replication and codes for several transcripts. Curr. Genet. **17:** 321–325.

Hanfler, J., H. Teepe, C. Weigel, V. Kruft, R. Lurz *et al.*, 1992 Circular extrachromosomal DNA codes for a surface protein in the (+) mating type of the zygomycete *Absidia glauca*. Curr. Genet. **22:** 319–325.

Hughes, J. E., and D. L. Welker, 1989 Copy number control and compatibility of nuclear plasmids in *Dictyostelium discoideum*. Plasmid **22:** 215–223.

Hughes, J. E., H. Ashktorab and D. L. Welker, 1988 Nuclear plasmids in the *Dictyostelium* slime molds. Dev. Genet. **9:** 495–504.

Hughes, J. E., G. J. Podgorski and D. L. Welker, 1992 Selection of *Dictyostelium discoideum* transformants and analysis of vector maintenance using live bacteria resistant to G418. Plasmid **28:** 46–60.

Hughes, J. E., H. Kiyosawa and D. L. Welker, 1994 Plasmid maintenance functions encoded on *Dictyostelium discoideum* nuclear plasmid Ddp1. Mol. Cell. Biol. **14:** 6117–6124.

Kiyosawa, H., J. E. Hughes, G. J. Podgorski and D. L. Welker, 1993 Small circular plasmids of the eukaryote *Dictyostelium purpureum* define two novel plasmid families. Plasmid **30:** 106–118.

Kiyosawa, H., J. E. Hughes and D. L. Welker, 1994 Compatible *Dictyostelium mucoroides* nuclear plasmids Dmp1 and Dmp2 both belong to the Ddp1 plasmid family. Plasmid **31:** 121–130.

Kiyosawa, H., J. E. Hughes and D. L. Welker, 1995 The replication origin position and its relationship to a negative trans-acting transcription regulator encoded by *Dictyostelium discoideum* nuclear plasmid Ddp1. Curr. Genet. **27:** 479–484.

Leiting, B., I. J. Lindner and A. A. Noegel, 1990 The extrachromosomal replication of *Dictyostelium* plasmid Ddp2 requires a *cis*-acting element and a plasmid-encoded *trans*-acting factor. Mol. Cell. Biol. **10:** 3727–3736.

Metz, B. A., T. E. Ward, D. L. Welker and K. L. Williams, 1983 Identification of an endogenous plasmid in *Dictyostelium discoideum*. EMBO J. **4:** 515–519.

Noegel, A., D. L. Welker, B. A. Metz and K. L. Williams, 1985 Presence of nuclear associated plasmids in the lower eukaryote *Dictyostelium discoideum*. J. Mol. Biol. **185:** 447–450.

Service, R. F., 1997 Amino acid alchemy transmutes sheets to coils. Science **277:** 179.

Shammat, I. M., C. M. Gonzales and D. L. Welker, 1998 *Dictyostelium discoideum* nuclear plasmid Ddp6 is a new member of the Ddp6 plasmid family. Genetics. (in press).

Slade, M. B., A. C. M. Chang and K. L. Williams, 1990 The sequence and organization of Ddp2, a high-copy-number nuclear plasmid of *Dictyostelium discoideum*. Plasmid **24:** 195–207.

Vincent, K. A., V. Ellison, S. A. Chow and P. O. Brown, 1993 Characterization of human immunodeficiency virus type 1 integrase expressed in *Escherichia coli* and analysis of variants with amino-terminal mutations. J. Virol. **67:** 425–437.

Yin, Y., and D. L. Welker, 1992 *Dictyostelium giganteum* plasmid Dgp1 is a member of the Ddp2 plasmid family. Plasmid **28:** 37–45.

Communicating editor: S. L. Allen