J Mol Evol (1996) 43:185–193

JOURNAL OF MOLECULAR EVOLUTION

Evolution of Promoter Sequences: Elements of a Canonical Promoter for Prespore Genes of *Dictyostelium*

Craig Miller,[1] John McDonald,[2] David Francis[2]

[1] Biology Department, University of Oregon, Eugene, Oregon 97403, USA
[2] Biology Department, University of Delaware, Newark DE 19716, USA

Received: 16 September 1995 / Accepted: 6 April 1996

Abstract. An attempt is made to define a minimal prespore promoter which contains all elements essential for correct regulation of expression of a prespore gene. The prespore genes of *Dictyostelium* are coregulated during development. Most begin transcription at the same early stage, and activity of all is restricted to prespore tissue during the later slug stage. Sequences 5′ to the coding sequences of eight prespore genes were searched for all elements proposed to control transcription and for new elements. The meaningfulness of occurrences of elements and pairs of elements in prespore promoters was evaluated by comparison with frequencies of occurrences in promoters of other, nonprespore genes. These comparisons resulted in definition of a canonical prespore promoter, a stretch of about 200 nucleotides containing at least one of each of three elements. Certain limitations were found on the spacing of elements. Orientation of elements with respect to each other appeared unrestricted. All elements often occurred in multiple copies. This structure suggests that individual copies of each element are not conserved during evolution, but instead continually appear and disappear.

Key words: *Dictyostelium* — Prespore genes — Regulatory sequences — Transcriptional elements — Promoter evolution

Correspondence to: D. Francis

Introduction

As a start toward defining the constraints on evolutionary change in regulatory DNA sequences, we searched for shared features of the promoters of a group of genes which have a similar pattern of expression. These are the prespore genes of *Dictyostelium,* a cellular slime mold.

Dictyostelium is a soil amoeba. The developmental process which leads to sporulation begins when starved, crowded amoebae aggregate by gathering in streams which flow to central mounds. The mound rapidly transforms into a wandering slug with two primary tissues. The rear four-fifths consists mainly of prespore amoebae, and the front one-fifth of prestalk cells. This arrangement may be maintained for many hours. The slug eventually builds a stalked fruiting body consisting of stalk cells derived from prestalk amoebae and spores from prespore amoebae.

Six prespore genes code for structural proteins of the spore coat and share a 13-amino-acid repeat sequence motif (Powell-Coffman and Firtel 1994; Yoder and Blumberg 1994). Presumably they have originated from a single ancestral gene. Four of these genes are very similarly expressed during development, mRNAs being first detectable in scattered cells of the central mound of aggregation and maximally concentrated in the prespore part of the slug. They are not expressed in prestalk cells, the boundary between the two tissues being very sharp. Three genes have slightly different patterns of expression: The Dp87 message first appears in isolated cells of aggregating streams (Ozaki et al. 1993), a few hours earlier than the others. Transcription of the PL3 message is uniquely dependent on multicellularity and stops

abruptly if mounds are disaggregated (Agarwal et al. 1994). At the slug stage, however, expression of all prespore genes is restricted to the same set of cells.

Two other prespore genes are unrelated to those coding for spore-coat proteins. These are pspA (D19), which may code for a cell adhesion protein, and D7, which is a unique sequence. Expression of these is very similar to that of the major group of spore-coat genes.

In general, expression of these genes is controlled by sequences in the promoter region, the region 5′ to the coding sequences. This region is mostly less than 1,000 bp long in *Dictyostelium* and consists of a set of regulatory elements, partly identified, embedded in context sequences of poorly understood function. This general description encompasses a range of specific possibilities for the minimal promoter which contains all elements required to give expression in prespore cells. At one extreme would be an invariant short cassette, consisting entirely of elements which bind regulatory factors. Selection would maintain the sequence of each element as well as such features as the number, order, orientation, and spacing of individual elements. An intermediate type of canonical promoter would be one in which the only invariant sequences are the regulatory elements; their number, order, orientation, spacing, and the sequence of the interelement context being free from selection. An extreme alternate would incorporate degenerate regulatory elements. In this case promoters of similarly regulated genes might have no recognizable sequence similarity.

We attempt to define a canonical prespore promoter by detailed comparisons of the sequences of those prespore promoters which have been subjected to deletion analyses. To help determine whether a proposed element is truly essential and specific to prespore promoters we have examined occurrence of elements in promoters of all *Dictyostelium* promoters and have calculated the probability of chance occurrence of elements. A regulatory element may work alone to regulate prespore-ness, in which case it is expected to occur in prespore promoters with a frequency much higher than expected by chance, or to occur in other promoters with a much-lower-than-chance frequency. Alternatively it may function only in combination with another element; and in this case its frequency of occurrence might be no higher than that expected from chance, but the combination of elements should be found only in prespore promoters. With this in mind, we have examined the patterns of occurrence of combinations of elements and of pairs and triplets of elements.

Results and Discussion

Possible Elements of Prespore Promoters

Several candidate elements have been proposed for the basic prespore promoter. The distribution of these in seven promoters is shown in Fig. 1 and given in detail in Table 1. Two kinds of evidence are examined in deciding the necessity of each element to the promoter. First, evidence from genetic and biochemical work supports or proves the function of an element in regulating prespore expression. Second, evidence from occurrences of an element or combination of elements can confirm whether it is specific to promoters of prespore genes or is found in all promoters.

The GRE

Most completely understood is the G-rich element (GRE). The version of known function consists of two copies of the heptamer DGKGKGD, separated by 4–7 bp (Powell-Coffman et al. 1994). It seems clear that this doublet element acts as a positive control element: deletion of a small promoter region containing it causes a drastic reduction in expression of the prespore gene pspA (D19) (Early and Williams 1989), without affecting timing or localization of expression. A protein (the G-box binding factor) which is active only in the later stages of development binds specifically to GRE doublets (Schnitzler et al. 1994). Point mutagenesis of a GRE doublet reduces both prespore expression of a reporter gene and binding affinity of the G-box binding factor to to a promoter containing this sequence (Haberstroh et al. 1991) The GRE appears to mediate the response to late-expressed genes to high continuous levels of external cAMP.

The GRE singlet heptamer is so degenerate that it may include several elements of different function. One version may mediate the transcriptional response to pulses of external cyclic AMP which is shown by several aggregation-stage genes. The Tai element (GGTGTGAT) (Tai and Podgoski, in prep.) and the Desbarats element (TGGTGTG) (Desbarats et al. 1992) appear to be two variants of this second cyclic AMP response element. The Tai variant, at least, appears to act as a singlet rather than as a doublet. GRE singlets might function in initiating expression of prespore genes at the time of aggregation in response to pulses of cyclic AMP, perhaps in combination with unidentified other elements.

Tables 1, 2, and 3 document the occurrences of GRE. GRE singlets occur in all classes of promoters at frequencies far in excess of that expected from nucleotide composition, which is 84% A+T for all promoters. GRE doublets, however, occur in prespore promoters at a frequency 30 times higher than expected from chance cooccurrences of GRE singlets (Table 3). But GRE doublets are not found in every prespore promoter, being clearly absent from the short sufficient promoter of pspA (D19). At least a singlet element appears to be an essential part of the canonical promoter.

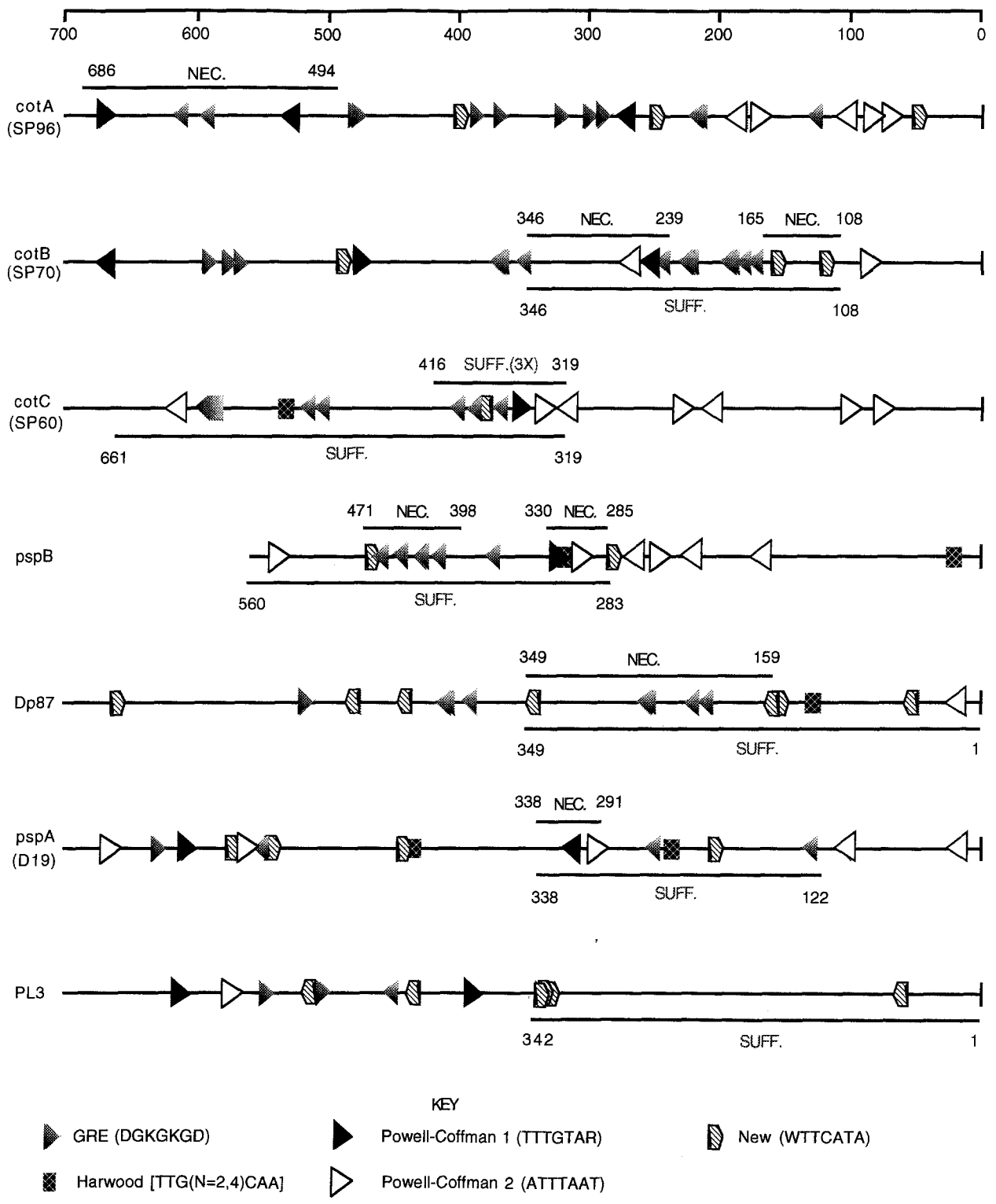

Fig. 1. Promoter elements of seven prespore genes which have been subjected to deletion analysis. In all cases 700 bp (or 560 bp in the case of pspB) is sufficient to drive prespore-specific expression of a receptor gene. Smaller sufficient or necessary regions are indicated. Complements of elements are indicated by *triangle pointing left*. Exact start points of each element are indicated in Table 1. The *SUFF.* region of PL3 in fact extends through codon 20 of the coding sequence, and the portion downstream of transcription start may contain regulatory elements (Blumberg, personal communication). Similarly the *SUFF.* region of Dp87 extends through codon 6. Data are from Tasaka et al. (1992) for cotA (SP96); Fosnaugh and Loomis (1993) for cotB (SP70); Ceccarelli et al. (1991) and Powell-Coffman et al. (1994) for cotC (SP60); Powell-Coffman and Firtel (1994) for pspB; Morio et al. (1994) for Dp87; Early and Williams (1989) for pspA (D19); and Yoder and Blumberg (1994) for PL3. No deletion analysis is available for an eighth gene, D7.

Table 1. Elements found in promoter sequences[a]

Class	<----------PRESPORE---------->								<--------PRESTALK-------->				<----------MISCELLANEOUS---------->				
Element	cotA (SP96)	cotB (SP70)	cotC (SP60)	pspB	Dp87	pspA (D19)	psvB (PL3)	D7	ecmA	ecmB	rasD	D11	CP2	PDE-A	PDE-V	PDE-L	UDPGP
GRE	1676	1444	598*	460*	1376*	632	551	691	1305*	1034*	162*	1341*	230	672	844*	1148*	838
DGKGKGD	1327	802*	590*	446*	520	554*	508	340*	1249*	816	151*	1043*	216	624*	766*	1140*	394*
	615*	594	517*	431*	412*	251*	454*	295*	1194*	808		701		462	138	1133*	
	594*	579	507*	418*	392*	131*	339	196	1111*	530		641		376		1093*	
	483	571	403*	374*	258*		331	168	1076	523		226		273		873*	
	390	373*	392*		221*			157	1050					150*		866*	
	371	351*	370*		211*			137*	828							788*	
	324	245*							816							776*	
	302	227*							616							759*	
	292	193*							484*							301	
	218*	183*							375*							142*	
	129*	175*														+174*	
																+185*	
Powell-Coffman 1	769	872*	750*	327	800*	609	1813*	727	1553*	1318*	137*	1215	718*	1155	1030	+416*	
TTTGTAR	673	761*	355		+67	314*	1750*	632	1025*	1214*		277*		858*	815		
	530*	669*	+24			+71	1468*	457	859*	1098*							
	274*	478	+37				1275			719							
		251*					709			509							
		+77*					616										
							392										
Powell-Coffman 2	924	1343*	1236*	542	970*	813	2396*	662	1600*	296*	107	948*		1140*	800*	1307	282
ATTTAQAT	186*	1338*	826	309	806*	783	1856	83	996*			580*		530	785	932	277
	175	1177*	616*	266*	18*	671	1677*	54	969			436*			641	919	192*
	104*	1113	336	247	+34*	568	941*	3*	666			157*			168	914*	
	84	269*	318*	220*		296	577		+55			137				813*	
	71	80	230	169*		105*						103				87	
		+96	206*	+112*		17*						83*				+246	
			102	+117*													
			78														
Harwood	2061		532	326	129	436	1342		1363	1241	299			234	737		
TTGNCAA	(TTGG)		(TT)	(TAGT)	(CT)	(AG)	(AAAC)		(TT)	(TA)	(GAAA)			(ATTT)	(AT)		
(N = 2 or 4)	1088					239				473	173			54	386		
	(GTTTT)					(ATAT)				(AA)	(TG)			(ATTTT)	(AC)		
	858									111					206		
	(AAAT)									(AT)					(AAAT)		
										337							
										(ATTTT)							
New	1430	951	381*	468	664	909*	2259	908	695			362*	564			347*	661
WTTCATA	1196*	946*		282	482*	575	1785*	707	399	491			442*	547	515*	399	
	249	159			343*	445	436*	98									
	50	122			162*	206											
					157												
					54*												
Total length	2246	1714	1421	781	1510	1093	2490	1160	1950	1717	700	1508	833	1162	1104	1860	1047
Length to *	2065	1461	1333	560	1380	1001	2408	1009	1696	1616	527	1479	739	1162	1104	1423	907
Accession #	S96842	M98546			D13973	X15980			X74046	M73676			M16039	M23449	M23449	M23449	M30467
References	15	13	5,19	14	1,11	3	18		39	16	10,17		35	20,25	20,25	20,25	21

The Harwood Element

While it may be that the essential feature of a prespore gene promoter is a positive regulatory element that is only activated in prespore cells, a very different possibility is that these promoters have a repressor element that responds to inhibitor signals in the prestalk zone and prevents expression there. All positive-acting elements might be common to prespore and prestalk promoters. To date the only known repressor element in *Dictyostelium* is the Harwood element (TTG $N_{2,4}$ CAA) (Harwood et al. 1993). The Harwood element occurs no more frequently in prespore promoters than expected by chance (Table 2). It is present in exact form in six of the eight prespore promoters examined, but not in any consistent combination with any other known element (Fig. 1, table 1). Its function in the promoter of the prestalk ecmB gene has been defined as restricting expression to a small subset of prestalk cells (Harwood et al. 1993). Point mutations in the element eliminate its function and cause more widespread expression of ecmB. No protein factor has yet been described which binds this element.

It is not known whether the Harwood element requires the near presence of a particular co-acting element to exert its effect, as is true of certain repressor elements in the promoters of *Drosophila* developmental genes (Gray et al. 1994). Different co-acting elements might occur in promoters of prespore and prestalk genes so that the repression would occur in response to different tissue-

Table 1. Extended

Class	<---------- MISCELLANEOUS ---------->																		
Element	carA-E	carA-L	a-man	BP74	gp24A	gp24B	gp80	dis-a	dis-g	gp1	gp2	v4a	v4b	v14	v18	13	h4	cell	1091
GRE DGKGKGD	538 505* 419*	386 315* 270* 257* 239* 213* +64* +603*	798 105* +45*	452* 275* 239* 172*	211 202	615* 199 192	679* 594* 515 499 153	225 186*		383* 303	85* 50*		471*		1714 826 238 69*			+142*	
Powell-Coffman 1 TTTGTAR	+351*	220* +233	987*	816 421*				130* 83*	104*	1638* 677*				689	1360*		440 401*		484
Powell-Coffman 2 ATTTAAT	961*	633 508* 119* +111*	29*	871* 396* 368* 217	32	957 513* 32	298 +117	253* +4* +51* 83*	407* 380* 369* 1573* +4*	1828 1703 1596* 316	531* 513	331 277*		1301*	1858* 1638 1308 1217* 1068 801* 201* 14	299*	11* 534	549 544 539 515 116 81*	440 182 +60*
Harwood TTGNCAA (N = 2 or 4)		+198 (GTCT)	530 (AAAA)	979 (GT) 611 (AT)		884 (TT)	914 (GT)		758 (AT) 212 (CCCC)	+71 (TT)				908 (GCCA)	1159 (AAAC) +33 (CCGT) +66 AAAC)			401 (AGGC)	
New WTTCATA			896*	679			866 612		+24	651	*			346	250* 113*	308		598*	
Total length	1294	1330	1176	1163	1071	1103	1323	613	1098	2029	1216	613	614	1186	2001	528	643	1345	615
Length to *	1010	695	1116	1128	1022	1054	1179	543	1035	1894	930	568	568	1147	1904	502	629	972	523
Accession #	L09637	L09637	M82822	M29237	M27588	M27588	X66583	J01282	M55332					X15383	X15382	L08391	X15388	M33861	M19467
Refe0rences	34	34	33	32	31	31	7	30	39	37	29	28	28	26,27	26,27		26,27		38

[a] Elements found in promoter sequences. Genes are listed across the top; cotA(SP96) through D7 are expressed only in prespore cells; ecmA through D11 are prestalk genes; the remainder are expressed at various times in vegetative or developing cells. pspA and D19 are two names for the same gene. Some other genes have been better known by the names of their proteins, which are given in parentheses.

The position of the 5′ nucleotide of each element is given, measuring upstream (5′ to) from the start of transcription (''start''). Elements with (+) numbers are on the 3′ side of transcription start. All available sequence upstream from the first codon was searched for each gene. Asterisks (*) indicate complements of elements.

Sequences for the D7 and D11 promoters were kindly provided by Daphne Blumberg. Published and Genbank sequences differed for disc A; the published sequence was used. The Harwood element at 173 in rasD is present in Louvion et al's (1991) sequence, but not that of Esch et al. (1992). Where multiple start-of-transcription sites were indicated in published material, numbering is from the farthest upstream site. The two transcripts of carA are clearly under regulation by different regions, so these are listed as two separate promoters. Similarly, three promoters are listed for pde. Minor corrections to published numbering were made for the promoters of D7, D11, pspA(D19), and CP2.

Degenerate nucleotides are indicated thus:

$$D = A, G, T;\ K = G, T;\ N = A, T, G, C;\ R = A, G;\ W = A, T$$

Key to references: *1*—Ozaki et al. (1993); *3*—Early and Williams (1989); *5*—Haberstroh et al. (1991); *7*—Desbarats et al. (1992); *10*—Esch et al. (1992); *11*—Morio et al. (1994); *13*—Fosnaugh and Loomis (1993); *14*—Powell-Coffman and Firtel (1994); *15*—Tasaka et al. (1992); *16*—Ceccarelli et al. (1991); *17*—Louvion et al. (1991); *18*—Yoder and Blumberg (1994); *19*—Haberstroh and Firtel (1990); *20*—Faure et al. (1990); *21*—Pavlovic et al. (1989); *25*—Podgorski et al. (1989); *26*—Singleton et al. (1989); *27*—Singleton et al. (1991); *28*—Mcpherson and Singleton (1993); *29*—Sucic et al. (1993b); *30*—Poole and Firtel (1984); *31*—Loomis and Fuller (1990); *32*—Hopkinson et al. (1989); *33*—Schatzle et al. (1992); *34*—Louis et al. (1993); *35*—Pears and Williams (1987); *37*—Sucic et al. (1993a); *38*—Giorda et al. (1989); *39*—Early et al. (1993)

specific signals. In this case it could act to repress prestalk expression when located within a prespore promoter. An alternative is that it responds to a specific signal localized in prestalk B cells. In this case its appearances in promoters of prespore genes might indeed be accidental and without function, since the signal would be absent from prespore tissue.

Other evidence suggests that the Harwood element may affect transcription of genes other than ecmB and at times other than the slug stage of development. In the

Table 2. Frequencies of elements in promoters of different gene classes[a]

	Prespore		Prestalk		Other		Expected
GRE	59 hits	0.48obsv	23	0.39	66	0.25	0.01
singlet	8/8 genes		4/4		18/24		
Harwood	9	.07	7	.12	19	.07	0.08
	6/8		3/4		12/24		
PC1	30	0.24	11	0.19	20	0.07	0.10
	8/8		4/4		14/24		
PC2	50	0.40	14	0.24	67	0.25	0.46
	8/8		4/4		21/24		
New	33	0.26	3	0.05	16	0.06	0.18
	8/8		2/4		12/24		

[a] The "obsv" frequency registers the frequency of occurrence per 100 bp of promoter sequence in all the genes of a class. For example, the GRE element occurs at 0.48 per 100 bp of prespore promoter on average (or 4.8 occurrences per kilobase). Elements occurring in both orientations were counted. The "hit" number is the total number of occurrences in all promoters of a class. The "genes" fraction is the fraction of genes of a category that have the element in their promoters. The "expected" frequency is calculated from the nucleotide sequences of the element and its complement and from the AT/GC content of promoter sequences (84% AT). Promoters of prespore genes total 12,415 bp, those of prestalk genes 5,875 bp, and those of other genes 26,967 bp. Note that all available promoter sequence was analyzed, not just the proximal 700 bp which is illustrated in Fig. 1.

Only nonoverlapping GRE elements were counted; that is, those whose start positions differed by seven or more nucleotides

Table 3. Frequencies of combinations of elements[a]

	Prespore		Prestalk		Other	
GRE doublet	15 hits	0.12 obsv	4	0.007	12	0.04
	7/8 genes	0.004 exp	3/4	0.002	6/24	0.001
Tight (PC1 + PC2)	6	0.05	2	0.03	4	0.015
= PC pair	5/8	0.004	2/4	0.002	4/24	0.001
Loose (PC1 + PC2)	17	0.14	4	0.07	9	0.03
	8/8	0.04	2/4	0.02	8/24	0.01
Loose ([loose PC1 + PC2]	6	0.05	2	0.03	2	0.007
+ GRE doublet)	6/8	0.02	2/4	0.0005	2/24	0.001
Loose (New + New)	6	0.05	0	0	0	0
	4/8	0.026	0	0.001	0	0.002
Loose	10	0.08	0	0	2	0.007
(New + GRE doublet)	7/8	0.015	0	0.002	2/24	0.001

[a] The "obsv" number underlined is the observed frequency of a combination in 100 bp. The "exp" value is the frequency of combinations expected from the observed frequencies of single occurrences given in Table 2, calculated using the Poisson distribution. The number of "hits" is the number of occurrences of a combination in all promoters of a category. The "genes" fraction indicates how many of the genes in a category have the combination in their promoter region.

A doublet of GRE elements has two elements wholly contained within 25 bp; spacing between the elements is 0 to 11 bp. A "tight" pair of PC1 + PC2 is contained in 37 bp. "Loose" combinations consist of elements wholly included in 114 bp.

The statistical significance of the association between New and GRE doublet elements in prespore promoters was tested using Monte Carlo simulations. Each simulated data set consisted of the observed set of promoters, each with its observed number of New and GRE doublet elements assigned positions at random. The random positions were constrained so that elements did not overlap. If there were fewer New elements within a promoter, the distance was measured for each New element to the nearest GRE doublet; if there were an equal number of each element type, or fewer GRE doublets, the distance was measured from each GRE doublet to the nearest New element. The cumulative frequency distribution of these nearest-neighbor distances was generated for 1,000 simulated data sets. The cumulative frequency distribution for the observed data fell outside 95% of the simulated distributions for a distance of 11–13 bp and 47–121 bp. This indicates that there are significantly more pairs of New with GRE doublet elements located 11–13 and 47–121 bp apart than would be expected if they were located randomly with respect to each other.

prestalk gene rasD, deletion of a 95-bp segment containing a Harwood element results in increased expression of a reporter gene in vegetative cells (Esch et al. 1992). Similarly, in the prespore gene Dp87, deletion of 65 bp containing a Harwood element results in derepression of expression during the vegetative stage (Mario et al. 1994). In no deletion study with cotC, pspB, Dp87, or pspA did removal of a piece containing a Harwood repressor element result in general expression of a prespore gene throughout the slug. In most studies, however, deletions were large, and might have eliminated closely associated repressor and activator elements together, with no detectable net effect on expression.

A variant Harwood element ($TTGN_3$ CAA) occurs in the first exon of PL3, and Blumberg (pers. comm.) has shown that this region acts as an inhibitor of prestalk expression, but only when adjacent to the PL3 promoter.

All this suggests that the Harwood-like element may

be important in some prespore promoters to restrict expression to prespore cells, but all this does not show that it is an essential component of the canonical prespore promoter.

The Powell-Coffman Pair

Two less thoroughly characterized elements which may function as a unit are the Powell-Coffman pair (PC1 and PC2) (TTTGTAR $(N)_{11,12}$ ATTTAAT). These have been identified in the necessary region of the minimal sufficient promoter of the prespore gene pspB (Powell-Coffman and Firtel 1994). Deletion of a small stretch which contains this pair from minimal sufficient promoters of pspA (D19), cotB (SP70), pspB, or cotC (SP60) erases prespore expression (Powell-Coffman et al. 1994; Early and Williams 1989; Fosnaugh and Loomis 1993; Powell-Coffman and Firtel 1994).

From this evidence, the PC1 and PC2 pair looks like a fair candidate for a prespore positive regulatory element. Unexpectedly, PC1 (TTTGTAR) occurs about as frequently in promoters of prestalk genes as in those of prespore genes (0.19 and 0.24, respectively; Table 2). PC2 (ATTTAAT) is nowhere more frequent than would be expected by chance. Powell-Coffman and Firtel (1994) noted that PC1 and PC2 elements in four prespore promoters occur together with an 11–12-bp spacing. To determine if close spacing is crucial, we searched for all PC1 and PC2 pairs with zero to 23 nucleotides between the half-elements. Such close pairs of PC1 and PC2 are far more frequent in all classes of promoters than expected from the frequencies of PC1 and PC2 alone (Table 3). Only five of the eight prespore promoters have a PC1 and PC2 pair. However, two of the three exceptions are Dp87 and PL3, which differ from the other genes in pattern of early expression (see above). Possibly the PC pair functions to activate prespore expression at the ordinary time during aggregation. If so, it may be a necessary part of the canonical prespore promoter.

The New Element

The above elements were suggested by deletion studies. There may well be other elements not identified as yet, since even the short sufficient promoters of cotB (SP70) and pspA (D19) have large stretches not occupied by known elements. These regions were surveyed for potential elements, using the COMPARE and DOTPLOT programs of the Wisconsin GCG package of programs. In addition to the known elements, a New element, WTTCATA, was discovered which also occurs in the minimal sufficient regions of other prespore promoters.

Evidence from deletion studies partly supports the New element as a functional entity. Firstly: a 238-bp minimal region sufficient for correct expression of a reporter gene was defined for cotB (SP70) (Fosnaugh and Loomis 1993), and the 3′ 58 bp of this region contains two New elements, and this piece is essential for expression. This is as expected if the New element is crucial for prespore expression. But in the cotC (SP60) gene, when a 25-bp piece containing the single New element was deleted from the 96-bp minimal sufficient region, prespore expression was *not* eliminated (Powell-Coffman et al. 1994). This result suggests that the New element is not essential to correct cotC (SP60) expression. However, the conclusion is clouded by the requirement of three tandem copies of the minimal cotC (SP60) promoter to give correct expression. A much longer 355-bp piece is required for correct expression when present as a single copy.

The New element occurs at *less* than its basic frequency in all except prespore promoters (Table 2). This alone suggests that it has meaning in the context of prespore promoters. It is not especially common as a doublet, but it is found far more frequently than expected loosely associated with GRE doublets (Table 3). The loose combination of New plus GRE doublet occurs in seven of the eight prespore promoters, and is found only twice elsewhere. The PC1 and 2 elements are not similarly associated with GREs.

In summary, the New element is significantly associated with the GRE doublet and the combination of New and GRE doublet occurs in most prespore promoters but rarely elsewhere. Available deletion results permit no clear conclusion about the necessity of the New element to prespore expression. Possibly its function is to maintain transcription specifically in prespore cells, once it has been initiated by signals working through other elements.

A Canonical Prespore Promoter

A loose association of GRE doublet, PC pair, and New element is common to most of these promoters. The combination of three elements is found in a 200-bp segment of the sufficient promoters of each of cotB (SP70), cotC (SP60), pspB, and pspA (D19). All these elements probably function in regulating transcription.

This picture of the canonical promoter falls between the first and second alternatives mentioned in the Introduction; elements are not so degenerate as to be unrecognizable, and only a few rules govern their pattern.

Not all prespore promoters contain the basic combination, Dp87 and PL3 being the notable exceptions. These two are exceptional in other respects. D87 is unusual in time of first expression (Ozaki et al. 1993). PL3 has a regulatory element in the first exon which is essential to expression in prespore cells (Blumberg, personal communication).

Degenerate but still functional versions of certain elements may occur in other nonstandard promoters. For example, PspA (D19) uniquely lacks a GRE doublet, but

it has a sequence which differs from a GRE element by one base (inverted: AGT-TGT, instead of AGTGTGT) at −146, and this is properly positioned to form a doublet with the GRE singlet at −131. If this combination is functional, it would provide this promoter with a loose (New plus GRE doublet) as well. Similarly, a degenerate version of PC2 (ATTA*T*AT instead of ATTAAAT) is located at −693 in the cotA promoter, where it could form the missing half of a tight PC1 +2 pair. If these identifications are correct, then all regular prespore promoters (still excepting Dp87 and PL3) have a PC1 + 2 pair, a GRE doublet, and a loose (New plus GRE) doublet.

Other features of the canonical promoter:

Number of Each Element

The number of GRE doublets varies from one to four in the entire promoter (except for PspA (D19), where there are isolated singlet elements). Similarly, there are up to seven New elements per promoter, and variable numbers of PC1 and PC2.

Orientation of Elements

The GREs of these promoters have a remarkable regional directionality; those within a 100–300-bp segment are oriented in the same direction. This feature may be an accidental result of whatever mutational mechanisms generate GREs, or it may be a selected feature. A corollary is that both elements of a GRE doublet always have the same orientation. The PC pair of elements mostly occurs with PC1 preceding PC2 as shown in the key to Fig. 1. The pair is oriented reversely in the cotB promoter, and the PC1 element is oppositely oriented with respect to PC2 in the pspA (D19) promoter. The small segment containing this unusual pair is necessary to the pspA (D19) minimal promoter, suggesting that it is functional. It is of interest that the pspA (D19) coding sequence is unrelated to that of the main group of prespore genes, implying that this promoter evolved from a separate ancestor, and perhaps explaining its unique version of the PC1 and 2 pair.

The New element occurs in both orientations, and indifferently upstream or downstream of the associated GRE pair with which it presumably cooperates.

Order of the Elements

There is no apparent restraint on the order of the GRE doublet, the PC pair, and the New element.

Spacing

A regularity in the spacing of New and GRE element was detected statistically (Table 3), and experimental work has shown that two unit GREs must be properly spaced to function as a binding site. Other restrictions on spacing of elements may exist, but not have been revealed by analysis.

It is probable that the essential set of elements only functions within an appropriate context which provides the spacing and physical structure necessary for transcription. About this context little can be said. It is extremely A/T rich, often containing unbroken runs of ten or more A's or T's. Regions containing such runs are crucial for expression of some genes in yeast (Iyer and Struhl 1995) and in *Dictyostelium* (Hori and Firtel 1994; Morio et al. 1994; Pavlovic et al. 1989). In addition, the interstitial material may contain sequences important for bending or kinking (Harrington and Winicov 1994; Hori and Firtel 1994) not easily recognized by sequence.

Conclusion

These features of the canonical promoter lead to the prediction that individual elements will show little conservation over evolution. This idea comes from three observations: (1) Some elements frequently occur in multiple copies within one promoter. Such redundancy suggests that individual copies of these elements can be lost without impairing promoter function. (2) The several promoters cannot be aligned so that individual elements of one can be identified with individual elements of another. This directly suggests that individual elements are not conserved. (3) The promoters of pspA (D19) and D7 have the canonical assembly of elements, yet the coding sequences of these genes are not related to other prespore genes. Mutation and selection appear necessary to explain this convergence, supposing that these two promoters evolved from ancestors with different assemblages of elements.

The hypotheses that similarities among the promoters are due to selection for a similar assembly of elements, and that elements are continually lost and continually generated by mutation, can be tested by observing spontaneous mutations in these promoters—for example, in repair-deficient strains (Welker and Deering 1978)—and by observing sequence variations in promoters of related *Dictyostelium* strains.

References

Agarwal A, Sloger MS, Oyama M, Blumberg DD (1994) Analysis of a novel cyclic AMP inducible prespore gene in *Dictyostelium discoideum:* evidence for different patterns of cAMP regulation. Differentiation 57:151–162

Ceccarelli A, Mahbubani H, Williams JG (1991) Positively and negatively acting signals regulating stalk cell and anterior-like cell differentiation in *Dictyostelium.* Cell 65:983–989

Desbarats L, Lam TY, Wong LM, Siu C-H (1992) Identification of a unique cAMP-response element in the gene encoding the cell adhesion molecule gp80 in *Dictyostelium discoideum.* J Biol Chem 267:19655–19664

Early AE, Williams JG (1989) Identification of sequences regulating the transcription of a *Dictyostelium* gene selectively expressed in prespore cells. Nucleic Acids Res 17:6473–6484

Early AE, Gaskell MJ, Traynor D, Williams JG (1993) Two distinct

populations of prestalk cells within the tip of the migratory *Dictyostelium* slug with differing fates at culmination. Development 118:353–362

Esch RK, Howard PK, Firtel RF (1992) Regulation of the *Dictyostelium* cAMP-induced, prestalk-specific Dd ras D gene: identification of cis-acting elements. Nucleic Acids Res 20:1325–1332

Faure M, Franke J, Hall AL, Podgorski GJ, Kessin RH (1990) The cyclic nucleotide phosphodiesterase gene of *Dictyostelium discoideum* contains three promoters specific for growth, aggregation, and late development. Mol Cell Biol 10:1921–1930

Fosnaugh KL, Loomis WF (1993) Enhancer regions responsible for temporal and cell-type specific expression of a spore coat gene in *Dictyostelium*. Dev Biol 157:38–48

Giorda R, Ohmachi T, Ennis HL (1989) Organization of a gene family developmentally regulated during *Dictyostelium discoideum* spore germination. J Mol Biol 205:63–69

Gray S, Szymanski P, Levine M (1994) Short-range repression permits multiple enhancers to function autonomously within a complex promoter. Genes Dev 8:1829–1838

Haberstroh L, Firtel RA (1990) A spatial gradient of expression of a cAMP-regulated prespore cell-type-specific gene in *Dictyostelium*. Genes Dev 4:596–612

Haberstroh L, Galíndo J, Firtel RA (1991) Developmental and spatial regulation of a *Dictyostelium* prespore gene: cis-acting elements and a cAMP-induced, developmentally regulated DNA binding activity. Development 113:947–958

Harrington RE, Winicov I (1994) New concepts in protein-DNA recognition: sequence-directed DNA bending and flexibility. Prog Nucleic Acid Res Mol Biol 47:196–270

Harwood AJ, Early A, Williams JG (1993) A repressor controls the timing and spatial localization of stalk cell-specific gene expression in *Dictyostelium*. Development 118:1041–1048

Hopkinson SB, Pollenz RS, Drummond I, Chisholm RL (1989) Expression and organization of BP74, a cyclic AMP regulated gene expressed during *Dictyostelium discoideum* development. Mol Cell Biol 9:4170–4178

Hori R, Firtel RA (1994) Identification and characterization of multiple A/T-rich cis-acting elements that control expression from *Dictyostelium actin* promoters: the *Dictyostelium actin* upstream activating sequence confers growth phase expression and has enhancer-like properties. Nucleic Acids Res 22:5099–5111

Iyer V, Struhl K (1995) Poly (dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. EMBO J 14:2570–2579

Loomis WF, Fuller DL (1990) A pair of tandemly repeated genes code for gp24, a putative adhesion protein of *Dictyostelium discoideum*. Proc Natl Acad Sci USA 87:886–890

Louis JM, Saxe CL III, Kimmel AR (1993) Two transmembrane signalling mechanisms control expression of the cAMP receptor gene CAR 1 during *Dictyostelium* development. Proc Natl Acad Sci USA 90:5969–5973

Louvion J-F, Scholder J-C, Pinaud S, Reymond CD (1991) Two independent promoters as well as 5′ untranslated regions regulate Dd *ras* expression in *Dictyostelium*. Nucleic Acids Res 19:6133–6138

Mcpherson CE, Singleton CK (1993) Nutrient-responsive promoter elements of the V4 gene of *Dictyostelium discoideum*. J Mol Biol 232:386–396

Morio T, Takeuchi I, Tasaka M (1994) Cooperation of positively and negatively acting promoter elements determines prespore-specific transcription of Dp87 gene in *Dictyostelium*. Mech Dev 45:59–72

Ozaki T, Nakao H, Orii H, Morio T, Takeuchi I, Tasaka M (1993) Developmental regulation of transcription of a novel prespore-specific gene (DP87) in *Dictyostelium discoideum*. Development 117:1299–1308

Pavlovic J, Haribabu B, Dottin RP (1989) Identification of a signal transduction response sequence element necessary for induction of a *Dictyostelium discoideum* gene by extracellular cyclic AMP. Mol Cell Biol 9:4660–4669

Pears CJ, Williams JG (1987) Identification of a DNA sequence element required for efficient expression of a developmentally regulated and cAMP-inducible gene of *Dictyostelium discoideum*. EMBO J 6:195–200

Podgorski GJ, Frank J, Faure M, Kessin RH (1989) The cyclic AMP phosphodiesterase gene of *Dictyostelium discoideum* utilizes alternate promoters and splicing for the synthesis of multiple mRNAs. Mol Cell Biol 9:3938–3950

Poole SJ, Firtel RA (1984) conserved structural features are found upstream from three co-ordinately regulated discoidin 1 genes of *Dictyostelium discoideum*. J Mol Biol 172:203–220

Powell-Coffman JA, Firtel RA (1994) Characterization of a novel *Dictyostelium discoideum* prespore-specific gene, PspB, reveals conserved regulatory sequences. Development 120:1601–1611

Powell-Coffman JA, Schnitzler GR, Firtel RA (1994) A GBF-binding site and a novel AT element define the minimal sequences sufficient to direct prespore-specific expression in *Dictyostelium discoideum*. Mol Cell Biol 14:5840–5849

Schatzle J, Bush J, Cardelli J (1992) Molecular cloning and characterization of the structural gene coding for the developmentally regulated lysozymal enzyme, alpha mannosidase, in *Dictyostelium discoideum*. J Biol Chem 267:4000–4007

Schnitzler GR, Fischer WH, Firtel RA (1994) Cloning and characterization of the G-box binding factor, an essential component of the developmental switch between early and late development in *Dictyostelium*. Genes Dev 8:502–514

Singleton CK, Manning SS, Ken R (1989) Primary structure and regulation of vegetative specific genes of *Dictyostelium discoideum*. Nucleic Acids Res 17:9679–9696

Singleton CK, Delude RL, Ken R, Manning SS, Mcpherson CE (1991) Structure, expression, and regulation of members of the developmentally controlled V and H gene classes from *Dictyostelium*. Dev Genet 12:88–97

Sucic JF, Luo S, Williamson BD, Yin Y, Rogers PV, Rutherford CL (1993a) Developmental and cAMP-mediated regulation of glycogen phosphorylase 1 in *Dictyostelium discoideum*. J Gen Microbiol 139:3043–3052

Sucic JF, Selmin O, Rutherford CL (1993b) Regulation of the *Dictyostelium* glycogen phosphorylase 2 gene by cyclic AMP. Dev Genet 14:313–322

Tai Y-S, Podgorski GJ (in prep)

Tasaka M, Hasegawa M, Nakata M, Orii H, Ozaki T, Takeuchi I (1992) Protein binding and DNase-1-hypersensitive sites in the cis-acting regulatory region of the spore-coat SP96 gene of *Dictyostelium*. Mech Dev 36:105–115

Welker DL, Deering RA (1978) Genetics of radiation sensitivity in the slime mould *Dictyostelium discoideum*. J Gen Microbiol 109:11–23

Yoder BK, Blumberg DD (1994) The promoter of a gene encoding a novel *Dictyostelium* spore coat protein. Dev Biol 163:38–48