Eur. J. Biochem. **265**, 441–448 (1999) © FEBS 1999

A *Dictyostelium* protein binds to distinct oligo(dA)·oligo(dT) DNA sequences in the C-module of the retrotransposable element DRE

Jens Horn[1], Anja Dietz-Schmidt[1], Ilse Zündorf[1], Jérôme Garin[2], Theodor Dingermann[1] and Thomas Winckler[1]

[1]*Institut für Pharmazeutische Biologie, Universität Frankfurt/Mainz (Biozentrum), Frankfurt, Germany;* [2]*Laboratoire de Chimie des Protéines, Grenoble, France*

The genome of the eukaryotic microbe *Dictyostelium discoideum* contains some 200 copies of the nonlong-terminal repeat retrotransposon DRE. Among several unique features of this retroelement, DRE is transcribed in both directions leading to the formation of partially overlapping plus strand and minus strand RNAs. The synthesis of minus strand RNAs is controlled by the C-module, a 134-bp DNA sequence located at the 3′-end of DRE. A nuclear protein (CMBF) binds to the C-module via interaction with two almost homopolymeric 24 bp oligo(dA)·oligo(dT) sequences. The DNA-binding drugs distamycin and netropsin, which bind to A·T-rich DNA sequences in the minor groove, competed efficiently for the binding of CMBF to the C-module. The CMBF-encoding gene, *cbfA*, was isolated and a DNA-binding domain was mapped to a 25-kDa C-terminal region of the protein. A peptide motif involved in the binding of A·T-rich DNA by high mobility group-I proteins (‘GRP’ box) was identified in the deduced CMBF protein sequence, and exchange of a consensus arginine residue for alanine within the CMBF GRP box abolished the interaction of CMBF with the C-module. The current data support the theory that CMBF binds to the C-module by detecting its long-range DNA conformation and interacting with A·T base pairs in the minor groove of oligo(dA)·oligo(dT) stretches.

Keywords: AT-rich DNA; *Dictyostelium*; distamycin; minor groove; retrotransposon.

Sequence-specific recognition of DNA by proteins is a central element in the regulation of cellular processes. Many DNA-binding proteins recognize particular base pair sequences in their target DNAs, hence performing ‘direct readouts’ of the information stored in the DNA sequences. In contrast, there are many proteins which, although binding specifically to certain DNA stretches, apparently recognize local DNA conformations rather than certain base pair sequences [1]. This ‘indirect readout’ is of special importance for the interaction of proteins with A·T-rich DNA. The conformations of oligo(dAdT)·oligo (dAdT) or oligo(dA)·oligo(dT) DNAs differ significantly from random DNA sequences. In oligo(dA)·oligo(dT) B-DNA fibres the A·T base pairs show high propeller twists resulting in maximized purine–purine stacking and the formation of non-Watson–Crick hydrogen bonds along the major grooves [2]. As a consequence, oligo(dA)·oligo(dT) B-DNA has a shorter average helical twist of 10.0 bp compared with random DNA (10.5 bp) and a narrow minor groove [2].

Although the base pairs in the minor groove of B-DNA offer only a few features for specific recognition by binding proteins compared with the major groove [3], the very special architecture of the narrow minor groove of A·T-rich DNA appears to be particularly important as an interaction partner for proteins binding to A·T-rich DNA. Several proteins have been isolated that interact with A·T-rich DNA primarily via the minor groove. Although no ubiquitous minor groove-binding protein domains have been found as yet, the arginine side chains in the ‘GRP’ motifs of mammalian high mobility group-I (HMG-I) proteins [4–6] and in other proteins containing GRP-like boxes [7–10] were shown to be critical for high-affinity recognition of A·T-rich DNA sequences.

Dictyostelium discoideum is a eukaryotic microbe which is being studied as a model system for the regulation of gene expression during developmental processes [11–14]. The genome of *D. discoideum* is remarkably A·T rich. Whereas the overall A·T content of exons in the *D. discoideum* genome is 65%, flanking sequences and introns averaged 80% and 87% A·T, respectively [15]. Marx *et al.* [16] pointed out that homopolymeric A·T tracts with a length of $N > 5$ occur more frequently in gene-flanking regions and introns than in coding regions, and that A·T tracts of $N > 10$ in gene-flanking regions occur in a regular phase that corresponds to the average nucleosome spacing, suggesting that A·T-homopolymeric DNA is largely excluded from the nucleosomes and is found in the nucleosomal linker regions. Oligo(dA)·oligo(dT) stretches have been shown to contribute to the regulation of the transcription rates by promoters in several organisms, e.g. humans [17], yeast [18–20] and *D. discoideum* [21–23].

The retroelement DRE in the *D. discoideum* genome is the first example reported to date of a nonlong-terminal repeat retrotransposon possessing great integration specificity upstream of transfer RNA genes [24,25]. A consensus full-length DRE element (DREa) encodes two ORFs enclosed by nonredundant UTRs termed the A-module and C-module [26]. Both the A-module and the C-module possess promoter activity and may be responsible for regulating the synthesis of DRE-specific plus strand and minus strand RNAs [27]. Transcription of minus strand RNAs starts within the oligo(A) stretch at the end of a DRE element (i.e. 3′ of the C-module).

Correspondence to T. Winckler, Institut für Pharmazeutische Biologie, Universität Frankfurt (Biozentrum), Marie-Curie-Strasse 9, D-60439 Frankfurt, Germany. Fax: + 49 69 7982 9662, Tel.: + 49 69 7982 9648, E-mail: Winckler@em.uni-frankfurt.de

Abbreviations: CMBF, C-module-binding factor; EMSA, electrophoretic mobility shift assay; HMG-I, high mobility group I proteins; TPR, tetratricopeptide.

(Received 22 June 1999, revised 30 July 1999, accepted 2 August 1999)

The heterogeneous population of minus strand RNAs is ≈ 2 kb in length [27] and contains neither extended ORFs nor consensus polyadenylation sites, suggesting that they may not be transported to the cytoplasm for protein translation.

We recently identified and purified a protein that binds to the C-module of genomic DRE copies *in vitro* [28]. The protein, which we termed C-module-binding factor (CMBF), was shown by footprinting analyses to bind at two ≈ 36-bp binding sites, which contain almost homopolymeric oligo(dA)·oligo (dT) cores of 24 bp length [28]. An isolated, synthetically produced CMBF-binding site (CMBS-2) retained its high affinity and selectivity of CMBF binding *in vitro* when inserted into a *D. discoideum* promoter that contained additional oligo(dAdT)·oligo(dAdT) and oligo(dA)·oligo(dT) stretches, indicating that the special conformation of CMBS-2 might be sufficient for a specific 'indirect readout' by CMBF in the highly A·T-rich *D. discoideum* genome [28]. Here we present the first detailed analysis of a *Dictyostelium* protein interacting with an almost homopolymeric oligo(dA)·oligo(dT) sequence. The interaction of CMBF with its binding sites in the C-module is achieved via minor groove binding within the oligo(dA)·oligo(dT) cores of the CMBF-binding sites by a critical arginine residue located within a GRP box at the C-terminus of the CMBF protein and by detection of the long-range conformation of the DNA within the C-module.

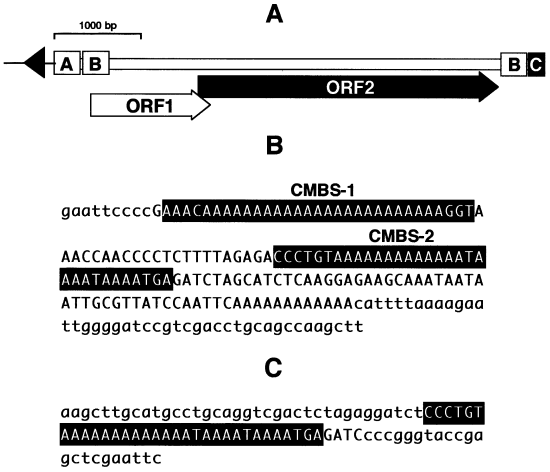

A
1000 bp
A
B
B
C
ORF1
ORF2
B
CMBS-1
gaattccccGAAACAAAAAAAAAAAAAAAAAAAAAAAAGGTA
CMBS-2
AACCAACCCCTCTTTTAGAGACCCTGTAAAAAAAAAAAAATA
AAATAAAATGAGATCTAGCATCTCAAGGAGAAGCAAATAATA
ATTGCGTTATCCAATTCAAAAAAAAAAAAcattttaaaagaa
ttggggatccgtcgacctgcagccaagctt
C
aagcttgcatgcctgcaggtcgactctagaggatctCCCTGT
AAAAAAAAAAAAATAAAATAAAATGAGATCcccgggtaccga
gctcgaattc

Fig. 1. Structure of DRE. (A) Schematic presentation of a full-length DREa element [26]. Two overlapping ORFs are flanked by untranslated 'modules'. The black triangle represents a DRE-asociated tRNA gene with its transcription orientation. (B) DNA sequence of the C-module (upper case letters) as cloned into the multiple cloning site of pUC9. The sequence shown is identical with the plus strand RNA. The DNA sequences protected by CMBF binding in footprinting experiments [28] are shown as black boxes. (C) DNA sequence of synthetically generated CMBS-2 inserted into the multiple cloning site of pUC19. Vector-encoded nucleotides are shown in lower case letters.

MATERIALS AND METHODS

Plasmids

The C-module of a DREa element was isolated from plasmid pUC9-C17 [28] as an *Eco*RI/*Hin*dIII fragment (Fig. 1B). CMBS-2 was generated synthetically [28] and ligated into the *Bam*H site of pUC19 (Fig. 1C).

Isolation and microsequencing of CMBF

CMBF was purified from vegetatively growing *D. discoideum* AX2 cells by using a scaled-up version of our recently described protocol [28]. Briefly, nuclei were prepared from 3.6×10^{11} cells (80 L of cell culture) and extracted with 400 mM KCl. Nuclear extract proteins were chromatographed on SP–Sepharose 4B, herring sperm DNA–Sepharose, MonoQ FPLC and MonoS FPLC columns. CMBF activity was monitored by its binding to the radiolabeled C-module [28]. CMBF peak fractions from the final MonoS column were pooled, dialyzed against 2 mM Tris/HCl (pH 8.0) and lyophilized. The dry powder (≈ 100 μg of total protein) was dissolved in 50 μL SDS/PAGE sample buffer [29] and proteins were separated in a semipreparative 10% SDS/PAGE [29]. The gel was stained with Coomassie Brilliant Blue [30], destained and the 115 kDa protein band was cut. The yield of 115 kDa CMBF protein was ≈ 30 μg. 'In gel' digestions of CMBF with the endoproteinases Lys-C and Asp-N, purification of peptides and peptide sequencing were performed as described previously [30]. The following peptides were obtained: Asp-N: Seq 6B, DLLGAPFPIYVFKQRPG; Seq 7B, XYIIKRGQPFVLTGTTQGWSR; Seq 8A, XXLIKAPNSNFIIMSNQPYQ; Seq 13A, XYLSHKLQPQYSYKSR; Seq 14B, DVHSKKRP(I)(V). Lys-C: Seq 67A, GGDVYNESRFIRPIDLLGAPFPIYVFK; Seq 70A, DFISYLQVSPEERNPK; Seq 71A, ELSSFYEIYK; Seq 71B, AIAYYTLRK; Seq 72B, DIAAFNYK; Seq 68A, RGQPFVLTGTTQG.

Isolation of the *cbfA* gene

Peptide Seq 7B was translated into a putative degenerate DNA sequence, and degenerate primers were designed to bind at the ends of both strands of the deduced DNA sequence [7B-forward, 5′-TA(CT)AT(CT)AT(CT)AAA(AC)-G(AT)GGTCAA-3′; 7B-reverse, 5′-(AT)C(GT)(AT)(GC)(AT) CCAACCTTG(AGT)GT-3′]. PCRs were performed using Amersham Taq polymerase and 200 ng genomic *D. discoideum* DNA (strain AX2) for 30 cycles at 95 °C for 1 min, 48 °C for 1 min and 72 °C for 1 min. PCR fragments were purified with the QIAspin PCR purification kit (Qiagen) and made radioactive by 1 min incubation with 3 U T4 DNA polymerase (New England Biolabs) in the absence of nucleotides, followed by fill-in reactions for 30 min in the presence of 330 μM dCTP, dGTP and dTTP, and 20 μCi [α^{32}P]dATP (3000 Ci·mmol^{-1}). Samples of radioactive PCR products were analyzed on 8% polyacrylamide/TBE gels. A radioactive DNA fragment of 60 bp was reproducibly generated (probe JH-7B). Genomic plasmid libraries were constructed from *D. discoideum* AX2 cells by ligating 25 μg genomic DNA digested with the appropriate restriction enzymes into the corresponding restriction sites of either pUC19 or pGEM7Zf(–) (Promega). The [^{32}P]-labeled probe JH-7B was used to screen a genomic *Eco*RI plasmid library, resulting in clone JH2.4 (nucleotides 432–2403 of the CMBF-coding region). Inverse PCR protocols were used to complete the 5′ DNA sequence of the *cbfA* gene. Genomic *D. discoideum* DNA (1 μg) was digested with *Bcl*I in 200 μL reaction volume. Aliquots of 1–5 μL of the digested DNA were subjected to intramolecular ligation in 500 μL reaction volumes using 4 U T4 DNA ligase (New England Biolabs). Between 1 and 10 μL of the ligation products were used in PCR reactions with primers CMBF2 and CMBF14 (CMBF2, 5′-GATGAATAAAGGTGATAAACC-3′; CMBF14, 5′-CGTTAAGGATATGATAGAATTC-3′). The resulting 1131 bp DNA fragment contained 578 bp of the putative promoter region of the *cbfA* gene. In order to isolate the 3′-end of the *cbfA* gene, a

*Kpn*I/*Sac*I genomic library was screened with the 1.7-kb DNA *Eco*RI insert of clone JH2.4 to obtain clone JH1.2, which carried 2529 bp of the *cbfA* gene and ≈ 500 bp of 3′-flanking sequence.

Electrophoretic mobility shift assays (EMSAs) and drug inhibition studies

CMBF was partially purified from growing *D. discoideum* AX2 cells as described previously [28]. 'DNA300' refers to a CMBF-containing fraction obtained by chromatography of nuclear extract proteins on SP–Sepharose and DNA–Sepharose [28]. Aliquots (0.5 µg) of the two nonspecific competitor DNAs poly(dAdT)·poly(dAdT) (Sigma #0883) and pUC9 were used in each binding reaction. Distamycin A and chromomycin A_3 were obtained from Sigma-Aldrich. Netropsin and mithramycin A were purchased from Fluka. The drugs were dissolved in 10 mM Tris/HCl (pH 7.1) and stored at −20 °C as 2 mM stock solutions. To study the ability of DNA minor groove-binding drugs to compete for the binding of CMBF to the C-module, the drugs were diluted in GP100 (20 mM Hepes-KOH, pH 7.9, 15% glycerol, 1 mM EDTA, 10 mM $MgCl_2$, 100 mM KCl, 1 mM dithiothreitol) to achieve the final concentrations used in the experiments. Competition experiments were performed by pre-incubating the radiolabeled probes and nonspecific competitor DNAs with DNA-binding drugs (0.02–20 µM) in GP100 for 10 min at ambient temperature prior to the addition of DNA-binding proteins for an additional 30 min. Quantitation of protein/DNA complexes was achieved by aligning the autoradiograms with the dried gels, cutting the retarded bands and counting their radioactivity in a liquid scintillation counter.

Expression of the CMBF DNA-binding domain in bacteria

The C-terminal 204 amino acids of CMBF were expressed in *Escherichia coli* cells. For this purpose nucleotides 3416–4027 of the CMBF-encoding gene (amino acids 1003–1206 of CMBF, $CMBF^{1003-1206}$) were amplified by PCR using plasmid JH1.2 as the template (primers: CMBF37, 5′-ACTGCCATGGAAGTGCATTCCAAGAAAAGACC-3′; CMBF40, 5′-CCCTCGAGTTTTTATTATTATAAAGTTCTTGTTTTAC-3′). The resulting DNA fragment was subcloned into pGEM-T (Promega) to yield pGEM-CMBF37/40. The DNA insert of pGEM-CMBF37/40 was isolated as a *Nco*I/*Not*I fragment and ligated into the corresponding sites of pET-22b(+) (Novagen). The resulting plasmid, pET-CMBF37/40, expressed a 28-kDa protein containing the pelB leader peptide fused to $CMBF^{1003-1206}$ and a hexahistidine tag. $CMBF^{1003-1206}$ was expressed in BL21(DE3) bacteria and partially purified by ion-exchange chromatography. Production of recombinant protein was induced with 1 mM isopropyl thio-β-D-galactoside (IPTG) for 4 h at 30 °C. One liter of cell culture was centrifuged and the bacteria were resuspended in 20 mL Hepes/NaOH (pH 7.2; buffer A) supplemented with 100 µg·mL^{-1} leupeptin and 1 mM $PhCH_2SO_2F$. The cells were lyzed by sonification and the lysate was centrifuged at 27 000 ***g*** for 30 min at 4 °C. The supernatant was applied at 4 °C to 5 mL of SP–Sepharose Fast Flow (Pharmacia) equilibrated in buffer A at 1 mL·min^{-1}. Bound proteins were eluted with successive steps of 100, 250, 500 and 1000 mM NaCl in buffer A. $CMBF^{1003-1206}$ eluted in the 500 mM NaCl fraction (SP500). The protein was adjusted to 50% glycerol and stored at −80 °C until further use. Recombinant $CMBF^{1003-1206}$ protein carrying a R1020A mutation was generated by site-directed mutagenesis of plasmid pET-CMBF37/40. The $CMBF^{1003-1206}$ (R1020A) protein was purified as described above. Protein concentrations were determined according to Bradford [31] using BSA as standard.

RESULTS

CMBF binding sites in the C-module

A full-length DREa element (Fig. 1A) contains in its 3′ UTR a 134-bp DNA sequence named the C-module. The C-module contains two almost homopolymeric oligo(dA)·oligo(dT) stretches of 24 bp length (Fig. 1B). In EMSAs CMBF was the only detectable DNA-binding activity capable of interacting with the C-module in a sequence-specific manner [28]. DNA footprinting analysis showed that CMBF bound to two ≈ 36-bp DNA motifs in the C-module, which consisted almost entirely of the 24 bp oligo(dA)·oligo(dT) tracts (CMBS-1 and CMBS-2, Fig. 1B). A third region of weak interaction (CMBS-3) reported in our previous study [28] was not reproducibly demonstrable in the footprinting experiments, and vectors containing a truncated version of the C-module containing only CMBS-3 could not compete for the binding by CMBF to a DNA

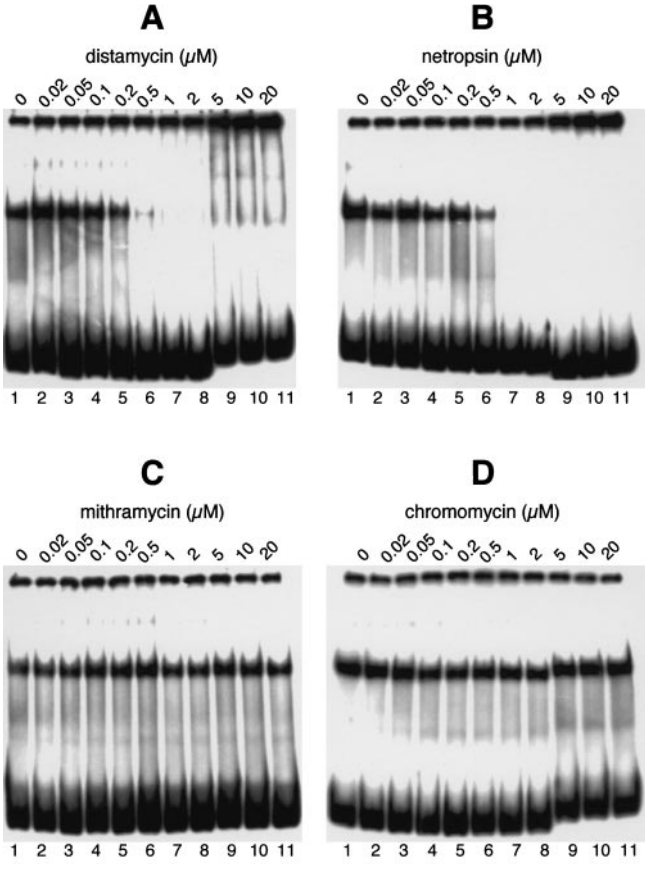


Fig. 2. Binding of CMBF to the C-module in the presence of minor groove-binding drugs. Binding of CMBF to the C-module was analyzed in EMSAs in the presence of distamycin (A), netropsin (B), mithramycin (C) or chromomycin (D). The drugs were diluted from 2 mM stock solutions in binding buffer GP100 and used at the concentrations indicated at the top of the figures. All mixtures contained 0.5 µg of each poly(dAdT)·poly(dAdT) and pUC9 as nonspecific competitor DNAs and the [^{32}P]-labeled C-module as probe. All samples were pre-incubated with the DNA-binding drugs prior to addition of CMBF (DNA300 fraction). Lane 1: no CMBF protein added.

fragment consisting of CMBS-1 and CMBS-2 (not shown). Our current interpretation is that CMBS-3 may represent a low-affinity CMBF binding site.

We set out to analyze in greater detail the interaction of CMBF with its A·T-rich target sequences in the C-module. Because of their high A·T content, the CMBF-binding sites were expected to be specific targets for high-affinity binding of DNA minor-groove-binding compounds such as distamycin and netropsin. These structurally related drugs bind to A·T-rich sequences in the minor groove, preferring oligo(dA)·oligo(dT) binding sites of at least 4 bp over alternating poly(dAdT)·poly(dAdT) sequences [32]. In contrast, mithramycin and chromomycin bind to the minor groove with preference for G·C base pairs [33,34]. We performed EMSAs in which the [^{32}P]-labeled C-module was incubated with the DNA-binding drugs prior to addition of partially purified CMBF. High-quality EMSAs with partially purified CMBF were only feasible in the presence of nonspecific competitor DNAs such as poly (dAdT)·poly(dAdT). Therefore the [^{32}P]-labeled C-module probe was mixed with 0.5 μg of poly(dAdT)·poly(dAdT) and pUC9 and pre-incubated with the DNA-binding drugs before adding CMBF. Distamycin and netropsin competed efficiently for the binding of CMBF to the C-module (Fig. 2A, B), whereas mithramycin and chromomycin displayed only minor effects (Fig. 2C, D). Competition by both distamycin and netropsin was complete at 1–2 μM. At low concentrations of distamycin (< 0.05 μM), CMBF binding to the C-module increased by ≈ 120% (Fig. 2A), while at higher concentrations it competed for the binding. This effect may result from some kind of 'optimization' of the ratio of nonspecific competitor DNA relative to CMBF concentration by binding of distamycin to poly(dAdT)·poly(dAdT). Quantitation of CMBF/DNA complexes revealed that weak competition ($< 20\%$) by both mithramycin and chromomycin occurred at a concentration of 1–5 μM (not shown). At concentrations exceeding 5 μM, nonspecific inhibition occurred with all four drugs, leading to altered migration of the unbound probe (Fig. 2). Even under these conditions, however, complexes of CMBF with the C-module were still observed in the presence of the G·C-specific minor groove binders (Fig. 2C, D). The quantitative data (not shown) further suggested that mithramycin was at least 10-fold less active as a competitor for CMBF binding to the C-module than the A·T-specific minor groove binders distamycin and netropsin.

As shown in our previous study [28], a synthetically generated CMBS-2 (36 bp) containing a d(TA$_{13}$TA$_4$TA$_4$T) kernel was recognized by purified CMBF when inserted into the A·T-rich background of a *D. discoideum actin15* promoter, whereas the *actin15* promoter alone was not a site for CMBF binding. This result suggested that CMBS-2 alone was sufficient for specific recognition by CMBF. Here we inserted CMBS-2 into the multiple cloning site of pUC19 (Fig. 1C) and used it as a 92-bp [^{32}P]-labeled *Eco*RI/*Hin*dIII fragment in EMSAs. CMBF bound to CMBS-2 albeit with somewhat lower affinity compared with the complete C-module (Fig. 3). Competition experiments using distamycin indicated that CMBF binding to this probe displayed C-module-like specificity (Fig. 3), confirming that interaction of CMBF with its binding sites in the C-module is primarily mediated by oligo(dA)·poly(dT) sequences.

Isolation of the CMBF-encoding gene

In order to evaluate the structural prerequisites of CMBF for specific interaction with A·T-rich sequences, we isolated the CMBF-encoding gene (*cbfA*). CMBF activity was purified from nuclear extracts by four successive chromatography steps [28]. From the final FPLC MonoS ion-exchange column CMBF activity eluted exactly with a 115-kDa protein visualized in silver-stained SDS/polyacrylamide gels [28]. The CMBF peak fractions from the MonoS gradients were pooled and subjected to a semipreparative SDS/PAGE. Approximately 30 μg of the 115 kDa protein were cut from the gel and hydrolyzed with endoproteinases. The N-terminal sequences of 11 peptides were obtained (see Materials and methods). The sequence of peptide 7B (20 amino acids) was selected to design degenerate primers specific for both strands of a putative DNA sequence deduced from the peptide amino acid sequence. The degenerate primers were used in PCRs with genomic *D. discoideum* DNA as a template. A 60-bp DNA fragment was reproducibly obtained and used as a radioactive probe to isolate the full-length *cbfA* gene from genomic *D. discoideum* libraries. The *cbfA* gene (GenBank accession number AF052006) codes for an ORF of 3627 bp, which is interrupted by a 409-bp intron starting at position +811 of the CMBF-encoding region (Fig. 4A). The intron boundaries were confirmed by comparing the genomic DNA sequence of the *cbfA* gene with the corresponding mRNA sequence obtained by PCR reactions of cDNA synthesized from RNA of vegetatively growing *D. discoideum* cells (not shown). The *cbfA* gene has the typical low G·C content found in all *D. discoideum* genes (26% and 7% G·C in the exons and intron, respectively). A consensus AATAAA polyadenylation signal [35] is present in the *cbfA* gene 17 bp upstream of the translation termination codon.

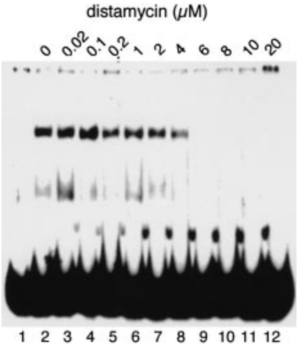
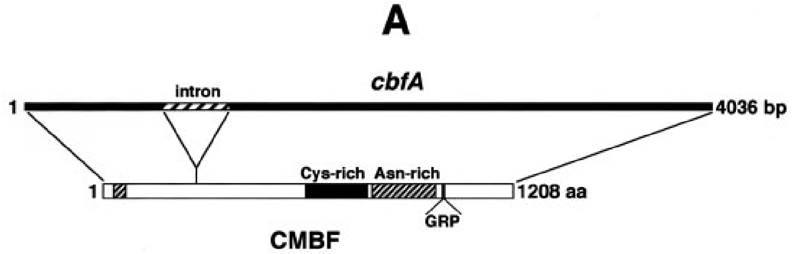


Fig. 3. Binding of CMBF to CMBS-2. EMSA showing the binding of CMBF (DNA300 fraction) to CMBS-2 subcloned in pUC19 (Fig. 1C). All mixtures contained 0.5 μg of each poly(dAdT)·poly(dAdT) and pUC9 as nonspecific competitor DNAs and the [^{32}P]-labeled CMBS-2 as probe. Samples were pre-incubated with the distamycin concentrations indicated at the top of the figure prior to addition of CMBF.

Characteristics of the deduced CMBF protein

The *cbfA* gene codes for a protein with 1208 amino acids and a calculated molecular mass of 138.8 kDa. All peptide sequences obtained from the purified 115-kDa CMBF protein were identified in the protein sequence deduced from the *cbfA* gene (Fig. 4B). However, the isolated CMBF protein was significantly shorter than expected from the *cbfA* gene sequence. Given that none of the determined peptide sequences were located in the N-terminal part of the deduced CMBF protein, the isolated 115-kDa CMBF protein may represent an N-terminal truncated version of CMBF due to *in vivo* processing of a CMBF precursor or proteolytic degradation of CMBF upon protein purification.



B

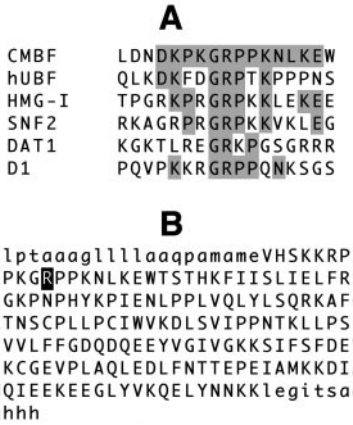
```
MSYTNTKEETQIINIENNDIKKRKFIENNNTNNNTNSNTNSNNKNNNNNNNNKSNNKKQN
NNNKKPRLKFTYDYNYKIISPFQKFLDYGICPTLNELYSEFYHILNENKCQKLKMELDEL
KKMNKGDKPIPIDFDPSKICHQHSKEDIEQIIFDPNDRHLIELAVRKFCGFLQKESNMVL
MEMIILSIGLLKSIEFPMYLPSVDNILNFNMEELIKAPNSNFIIMSNQPYQTTSPEIVED
YIIKRGQPFVLTGTTQGWSRSNMFTLDFLSERYSEMELINSPRNNETHTDLQGWRMKDFI
SYLQVSPEERNPKHLYGKDIACPREWQEYLSHKLQPQYSYKSRFDLVSHLPDYLQPETLL
VYIGSNGTYTPGHIDMCGSLSQNLMVSSDQDAFAWWFIVPTEYKDEALKFWGDKGGDVYN
ESRFIRPIDLLGAPFPIYVFKQRPGDFIFVPPDSVHQVVNCGPGISTKVAWNSISLKSLP
ISYFSSLPHTRRMAKPELFRIKAIAYYTLRKIMGDVENTNFNTIDVNDVIDIIAPLLEIF
HNILQTESILIPKPNYPYCNGETIPFLQPFKYFNGDRIQDRRCDHCNSDIFNRCYHCKPM
MVKVRISVSIVSRLVSVVNSISRLWYSKNSSLIANLRKKELSSFYEIYKNLLAHSGRRPK
EVDDIITKSTDRVSDECGFLTTATVAYHVVFYSSQKKIKCHRCEKRFKKFSIIFCTNCNA
RFCEQCVVNTFGQNFQVLMKRNEWECFCCKGLCDCSNCTSNSNSSNHPRILNNNQQLGLP
YNNNNNSNNNNNNNINNNNNNNNNNINNNNNNMNNNNSINNNNNNNNNNNINNNNINNNN
HHNNNGNNNLNSSYSSLNALSSLSQQQSYGSYDNYNNNNNNNNYNNNNNNNGHIQILKSG
RQYDDEQSSSSGSGSSNSTPTKPRPRNGGDDGLMSHFSGNNNNNNNHHNNNNNNNNNNHHM
MSHHHHNNNNNNNNNNNPTTSSLSSLSTSLSSSSTSTQKPMDVHSKKRPIVLDNDKPKGR
PPKNLKEWTSTHKFIISLIELFRSSNNAILGKPNPHYKPIENLPPLVQLYLSQRKAFGGV
LWAKTNSCPLLPCIWVKDLSVIPPNTKLLPSLIQGKKIVVLFFGDQDQEEYVGIVGKKSI
FSFDEVNQTLLLKCGEVPLAQLEDLFNTTEPEIAMKKDIAAFNYKNQIEEKEEGLYVKQE
LYNNKKII
```

Fig. 4. Primary structure of CMBF.
(A) Schematic drawing of the *cbfA* gene. In the deduced CMBF protein the locations of the cysteine-rich domain (black box) and the poly(asparagine) stretches (hatched boxes) are indicated. 'GRP' refers to the location of the GRP box DNA-binding motif critical for the DNA-binding activity of CMBF. (B) Amino acid sequence of the CMBF protein as deduced from the *cbfA* gene. The cysteine-rich sequences are shown in black boxes. The gray boxes highlight the positions of the peptides whose sequences derived from the purified 115-kDa CMBF protein.

Figure 4B shows the amino acid sequence of CMBF as deduced from the *cbfA* gene sequence. CMBF contains three cysteine-rich motifs situated in the central part of the protein. Computer searches [36] showed similarities of one part of the cysteine-rich motif ($CXHCX_7CXHC$, Fig. 4B) with $CX_2CX_7CX_2C$ motifs present in several proteins with as yet unidentified functions that have recently emerged from several genome sequencing projects. The central part of the CMBF zinc-finger-like structure, $HX_3YX_7CHXCX_{11}CX_2CX_4CX_2C$, was found to be related to a group of human zinc finger proteins with as yet unknown functions whose sequences were identified by the Human Genome Project. It is noteworthy that a histidine residue, which is conserved in the human zinc finger proteins ($HX_3HX_7CX_2C$), is replaced by a tyrosine residue in CMBF ($HX_3YX_7CX_2C$). Similarity of CMBF to proteins in the GenBank and SwissProt data bases resides in a 154 amino acid sequence (319–472 in CMBF) that exhibits significant similarities ($P = 1 \times 10^{-5}$) to a group of functionally diverse proteins which share 9–11 tetratricopeptide (TPR) motifs [37]. However, the region of similarity of these proteins and CMBF is not in the TPR repeats and its function is currently unknown.

Identification of a DNA-binding domain in the CMBF protein

In the CMBF protein an ≈ 200 amino acid asparagine-rich region separates the zinc-finger-like domain from a ≈ 25-kDa C-terminal domain (Fig. 4B). Close inspection of the amino acid sequence of this domain revealed a sequence motif similar to GRP boxes of other A·T-rich DNA-binding proteins (Fig. 5A). In order to evaluate the contribution of the GRP box motif in CMBF to its DNA-binding activity we expressed the C-terminal domain of CMBF ($CMBF^{1003–1206}$) in bacteria. $CMBF^{1003–1206}$ was expressed as a fusion with the bacterial pelB leader peptide at the N-terminus and a hexahistidine tag at the C-terminus (Fig. 5B). A high level of $CMBF^{1003–1206}$ protein was produced in both the soluble and particulate fractions of bacteria carrying the pET-CMBF37/40 plasmid. $CMBF^{1003–1206}$ was partially purified from high-speed supernatants of bacterial lysates by a single ion-exchange chromatography step, taking advantage of the basic nature of the $CMBF^{1003–1206}$ protein. The time required to separate $CMBF^{1003–1206}$ from lysate proteins proved crucial to prevention of proteolytic cleavage and generation of an N-terminal truncated $CMBF^{1003–1206}$ variant that was unable to bind to the C-module. This rapid separation was best achieved using SP–Sepharose chromatography instead of Ni^{2+}NTA affinity chromatography. As shown in Fig. 6A (lane 3) the recombinant $CMBF^{1003–1206}$ protein was enriched in the 500 mM NaCl column eluate (SP500) of lysates prepared from bacteria transformed with pET-CMBF37/40. In contrast, no protein of comparable size was enriched in the corresponding SP500 fraction prepared from mock-transformed bacteria (Fig. 6A, lane 2). EMSAs using the recombinant $CMBF^{1003–1206}$ protein and the radiolabeled C-module showed the formation of specific protein/DNA complexes (Fig. 6B, lanes 5–7), which were absent in fractions prepared under identical conditions from lysates of mock-transformed bacteria (Fig. 6B, lanes 2–4). The specificity of binding to A·T-rich kernels of the C-module was apparently conserved in the $CMBF^{1003–1206}$ protein, because its binding to DNA was efficiently competed



```
CMBF     LDNDKPKGRPPKNLKEW
hUBF     QLKDKFDGRPTKPPPNS
HMG-I    TPGRKPRGRPKKLEKEE
SNF2     RKAGRPRGRPKKVKLEG
DAT1     KGKTLREGRKPGSGRRR
D1       PQVPKKRGRPPQNKSGS
```

```
mkyllptaaaglllllaaqpamameVHSKKRPIVL
DNDKPKGRPPKNLKEWTSTHKFIISLIELFRSSN
NAILGKPNPHYKPIENLPPLVQLYLSQRKAFGGV
LWAKTNSCPLLPCIWVKDLSVIPPNTKLLPSLIQ
GKKIVVLFFGDQDQEEYVGIVGKKSIFSFDEVNQ
TLLLKCGEVPLAQLEDLFNTTEPEIAMKKDIAAF
NYKNQIEEKEEGLYVKQELYNNKKlegitsaaal
ehhhhhh
```

Fig. 5. Expression of the C-terminal 25-kDa fragment of CMBF (CMBF$^{1003-1206}$) in bacteria. (A) Alignment of GRP boxes from CMBF, human UBF [8], human HMG-I [38], yeast SNF2 [39], yeast DAT1 [9] and *Drosophila* D1 [7]. Only one representative GRP box is shown for each aligned protein. Note that part of the sequence given for HMG-I (TPKRPRGRPKK) represents the 'A·T-hook' [5]. (B) Amino acid sequence of CMBF$^{1003-1206}$ expressed in bacteria carrying the plasmid pET-CMBF37/40. Vector-borne amino acids are shown in lower case letters. The arginine-1020 mutated in the CMBF$^{1003-1206}$(R1020A) variant is shown in the black box.

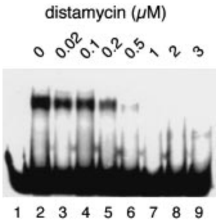


Fig. 7. Binding of recombinant CMBF$^{1003-1206}$ to CMBS-2. EMSA showing the binding of CMBF$^{1003-1206}$ (SP500 fraction) to CMBS-2 subcloned in pUC19 (Fig. 1C). All mixtures contained 0.1 μg of poly(dAdT)·poly(dAdT) as nonspecific competitor DNA and the [^{32}P]-labeled CMBS-2 as probe. Samples were pre-incubated with the distamycin concentrations indicated at the top of the figure prior to addition of CMBF$^{1003-1206}$. Lane 1: no protein added.

by distamycin (Fig. 6B, lanes 8–10), and CMBF$^{1003-1206}$ was able to bind to the CMBS-2 probe similar to authentic CMBF (Fig. 7).

In order to evaluate the role of the arginine residue in the GRP box in DNA binding by CMBF, we replaced arginine-1020 of CMBF with alanine by means of site-directed mutagenesis. The CMBF$^{1003-1206}$ (R1020A) mutant was expressed in bacteria and the recombinant protein was purified by ion-exchange chromatography (Fig. 6A, lane 4). As shown in Fig. 6B (lanes 11–13), the mutant CMBF$^{1003-1206}$ (R1020A) protein was unable to bind to the C-module probe, suggesting that arginine-1020 was critical for DNA binding by CMBF.

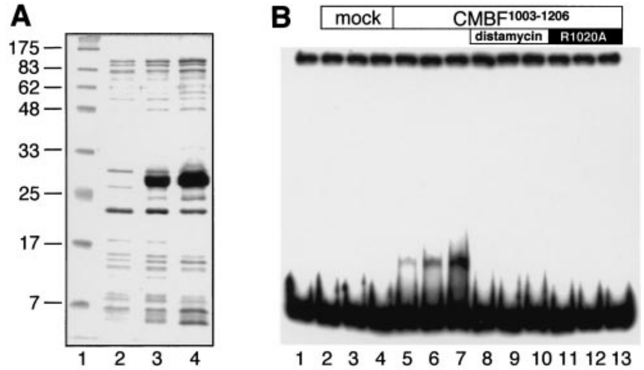


Fig. 6. Identification of a DNA-binding domain in CMBF. (A) Coomassie Brilliant Blue-stained SDS/polyacrylamide gel of the SP500 fractions derived from SP–Sepharose chromatographies performed with extracts of cells transformed with pET-22b (lane 2), pET-CMBF37/40 (lane 3) and pET-CMBF37/40(R1020A) (lane 4). Lane 1 shows marker proteins whose size is given at the left side of the figure. (B) EMSA using the SP500 fractions prepared from cells transformed with pET-22b (lanes 2–4), pET-CMBF37/40 (lanes 5–10) and pET-CMBF37/40(R1020A) (lanes 11–13). All mixtures contained 0.5 μg of each poly(dAdT)·poly(dAdT) and pUC9 as nonspecific competitor DNAs and the [^{32}P]-labeled C-module as probe. Approximately 10, 20 and 50 ng of protein were incubated with [^{32}P]-labeled C-module as probe in lanes 2–4, 5–7, 8–10 and 11–13 (lane 1: no protein added). In lanes 8–10 distamycin was added at 2 μM concentration.

DISCUSSION

CMBF binds to A·T-rich DNA sequences in the C-module

This report is the first detailed analysis of a *D. discoideum* DNA-binding protein displaying sequence-specific recognition of oligo(dA)·oligo(dT) sequence motifs, and to our knowledge the first description of a GRP box, 'A·T-hook'-like DNA-binding motif in *D. discoideum*. CMBF was first discovered through its specific interaction with the DRE C-module *in vitro*. We have shown that the oligo(dA)·oligo(dT) core of CMBS-2 within the C-module is important for the interaction with CMBF. Selective recognition of a DNA sequence by a binding protein is determined by the local conformation of the DNA molecule, which in turn is dictated by the nucleotide sequence [1]. We used DNA minor groove-binding drugs capable of discriminating A·T from G·C base pairs to support our previous data that CMBF interacts with the C-module via the oligo(dA)·oligo(dT) kernels of the individual binding sites. The competition experiments shown in Fig. 2 suggest that CMBF binding to the C-module is primarily achieved via the minor groove. However, we cannot exclude the possibility that distamycin and netropsin, even at low concentrations, induced subtle structural alterations in the C-module conformation leading to loss of CMBF binding. As discussed below, the presence of a GRP box motif that proved crucial for DNA binding by CMBF suggests that both recognition of the long-term DNA conformation of the C-module and interaction within the minor groove of oligo(dA)·oligo(dT) are equally important for specific and high-affinity DNA binding by CMBF. It is not yet clear what determines the specificity of CMBF for its binding sites in the C-module *in vivo* given that the highly A·T-enriched *D. discoideum* genome contains multiple oligo(dA)·oligo(dT) stretches. The physiological relevance of the CMBF DNA-binding specificity observed *in vitro* remains to be determined. Although we have shown that CMBF is capable of recognizing CMBS-2 in the A·T-rich backgound of a *D. discoideum actin15* promoter *in vitro* [28], the affinity and specificity of CMBF for oligo(dA)·oligo(dT) DNA may be influenced by chromatin structures *in vivo*.

CMBF contains a GRP box which contributes to DNA binding

Extensive purification of the C-module binding activity revealed a 115-kDa protein as the source of C-module-binding activity in nuclear extracts. The CMBF-encoding gene, *cbfA*, has coding capacity for a protein significantly larger than the purified CMBF protein. Evidence was presented to show that the cloned *cbfA* gene encodes CMBF. First, all peptides isolated from purified CMBF were found in the amino acid sequence deduced from the *cbfA* gene. Secondly, a monoclonal antibody raised against CMBF$^{1003-1206}$ detected a 115-kDa protein, but no 139 kDa protein, in nuclear extracts of *D. discoideum* cells (data not presented). Thirdly, and most importantly, the recombinant CMBF$^{1003-1206}$ protein bound the C-module probe in EMSAs was very similar to the CMBF protein purified from *D. discoideum* cells.

HMG-I proteins contain 'A·T-hooks' as represented in the peptide sequence TPKRPRGRPKK ([5]; Fig. 5A). HMG-I binds to any six or more consecutive A or T tracts with nanomolar affinities, and the 'A·T-hook' has been shown to exert structural features similar to netropsin and Hoechst 33258 [5]. GRP boxes similar to 'A·T-hooks' are found in the *DAT1* gene product in *Saccharomyces cerevisiae*, in the D1 protein of *Drosophila melanogaster* and in human UBF (Fig. 5A). Mutagenesis of the arginine within the GRP motifs of the DAT1 protein was shown to reduce high-affinity DNA binding [9]. We have shown by mutation analysis that CMBF contains an arginine residue within a single GRP box-like motif which contributes to DNA binding. We can not currently exclude that sequence motifs other than the GRP box motif contribute to the high-affinity C-module binding by CMBF. In particular we do not yet know the function of the zinc-finger-like sequence in the CMBF protein, which may either contribute to DNA binding across the flexible poly(asparagine) spacer separating it from the GRP box, or represent a structural component of CMBF which stabilizes its overall conformation.

Does CMBF regulate DRE transcription?

The hybridization of DRE plus strand and minus strand RNAs may down-regulate the level of translatable DRE plus strand RNA in the cell. If so, the minus strand synthesis rate would be a powerful regulatory checkpoint at which the host cell could interfere with the transposition activity of DRE elements. CMBF as a DNA-binding protein specifically binding the C-module is an excellent candidate to represent a host transcription factor regulating DRE expression. There is some evidence that interaction with DRE elements is not the only cellular function of CMBF. We know from Southern blot analysis of genomic DNA derived from various *Dictyostelium* species that, although DRE sequences are only detectable in *D. discoideum*, all *Dictyostelium* species analyzed to date express CMBF activity (T. Winckler, unpublished data). This could mean that CMBF is a DNA-binding protein involved in cellular functions other than regulating DRE expression or transposition. Gene disruption by homologous recombination is a convenient method in the analysis of *D. discoideum* gene function. Efforts to 'knock-out' *cbfA* have so far failed. One possible explanation for this would be that *cbfA* expression is essential for growth of *D. discoideum* cells, but the function of CMBF in *D. discoideum* cells remains to be determined. Because oligo(dA)·oligo(dT) stretches occur at high frequencies in the A·T-rich *D. discoideum* genome, the characterization of A·T-DNA specific DNA-binding proteins may provide important insights into how *D. discoideum* proteins regulate gene expression via modulation of promoter strengths at oligo(dA)·oligo(dT) stretches. The analysis of CMBF described in this report represents a further step towards a deeper understanding of gene regulation in *D. discoideum.*

ACKNOWLEDGEMENTS

We thank R. Marschalek for helpful discussions and J. B. Dressman for reading the manuscript. We are grateful to D. Pätsch and C. Tschepke for expert technical assistance. This work was supported by grants from the Deutsche Forschungsgemeinschaft to T. W. and T. D. (DI 268/7-1 and DI268/7-2).

REFERENCES

1. Travers, A.A. (1989) DNA conformation and protein binding. *Annu. Rev. Biochem.* **58**, 427–452.
2. Nelson, H.C.M., Finch, J.T., Luisi, B.F. & Klug, A. (1987) The structure of an oligo(dA) oligo(dT) tract and its biological implications. *Nature* **330**, 221–226.
3. Seeman, N.C., Rosenberg, J.M. & Rich, A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl Acad. Sci. USA* **73**, 804–808.
4. Solomon, M.J., Strauss, F. & Varshavsky, A. (1986) A mammalian high mobility group protein recognizes any stretch of six AT base pairs in duplex DNA. *Proc. Natl Acad. Sci. USA* **83**, 1276–1280.
5. Reeves, R. & Nissen, M.S. (1990) The AT–DNA-binding domain of mammalian high mobility group I chromosomal proteins. *J. Biol. Chem.* **265**, 8573–8582.
6. Huth, J.R., Bewley, C.A., Nissen, M.S., Evans, J.N.S., Reeves, R., Gronenborn, A.M. & Clore, G.M. (1997) The solution structure of an HMG-I(Y)–DNA complex defines a new architectural minor groove binding motif. *Nat. Struct. Biol.* **4**, 657–665.
7. Ashley, C.T., Pendleton, C.G., Jennings, W.W., Saxena, A. & Glover, C.V. (1989) Isolation and sequencing of cDNA clones encoding *Drosophila* chromosomal protein D1. A repeating motif in proteins which recognize at DNA. *J. Biol. Chem.* **264**, 8394–8401.
8. Jantzen, H.M., Admon, A., Bell, S.P. & Tjian, R. (1990) Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins. *Nature* **344**, 830–836.
9. Reardon, B.J., Winters, R.S., Gordon, D. & Winter, E. (1993) A peptide motif that recognizes AT tracts in DNA. *Proc. Natl Acad. Sci. USA* **90**, 11327–11331.
10. Reardon, B.J., Gordon, D., Ballard, M.J. & Winter, E. (1995) DNA binding properties of the *Saccharomyces cerevisiae DAT1* gene product. *Nucleic Acids Res.* **23**, 4900–4906.
11. Gross, J.D. (1994) Developmental decisions in *Dictyostelium discoideum. Microbiol. Rev.* **58**, 330–351.
12. Firtel, R.A. (1995) Integration of signaling information in controlling cell-fate decisions in *Dictyostelium. Genes Dev.* **9**, 1427–1444.
13. Loomis, W.F. (1996) Genetic networks that regulate development in *Dictyostelium* cells. *Microbiol. Rev.* **60**, 135.
14. Schaap, P., Tang, Y.H. & Othmer, H.G. (1996) A model for pattern formation in *Dictyostelium discoideum. Differentiation* **60**, 1–16.
15. Marx, K.A., Hess, S.T. & Blake, R.D. (1993) Characteriztics of the large (dA) (dT) homopolymer tracts in *D. discoideum* gene flanking and intron sequences. *J. Biomol. Struct. Dyn.* **11**, 57–66.
16. Marx, K.A., Hess, S.T. & Blake, R.D. (1994) Alignment of (dA). (dT) homopolymer tracts in gene flanking sequences suggests nucleosomal periodicity in *D. discoideum* DNA. *J. Biomol. Struct. Dyn.* **12**, 235–246.
17. Fashena, S., Reeves, R. & Ruddle, N.H. (1992) A poly(dA–dT) upstream activating sequence binds high-mobility group I protein and contributes to lymphotoxin (tumor necrosis factor-β) gene regulation. *Mol. Cell. Biol.* **12**, 894–903.
18. Russel, D.W., Smith, M., Cox, D., Williamson, V.M. & Young, E.T. (1983) DNA sequences of two yeast promoter-up mutants. *Nature* **304**, 652–654.
19. Struhl, K. (1985) Naturally occurring poly(dA–dT) sequences are

upstream promoter elements for constitutive transcription in yeast. *Proc. Natl Acad. Sci. USA* **82**, 8419–8423.

20. Lue, N.F., Buchman, A.R. & Kornberg, R.D. (1989) Activation of yeast RNA polymerase II transcription by a thymidine-rich upstream element *in vitro*. *Proc. Natl Acad. Sci. USA* **86**, 486–490.

21. Pavlovic, J., Haribabu, B. & Dottin, R.P. (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.

22. Hori, R. & Firtel, R.A. (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.

23. Ramalingam, R., Blume, J.E., Ganguly, K. & Ennis, H.L. (1995) AT-rich upstream sequence elements regulate spore germination-specific expression of the *Dictyostelium discoideum celA* gene. *Nucleic Acids Res.* **23**, 3018–3025.

24. Marschalek, R., Brechner, T., Amon-Böhm, E. & Dingermann, T. (1989) Transfer RNA genes: landmarks for integration of mobile genetic elements in *Dictyostelium discoideum*. *Science* **244**, 1493–1496.

25. Hofmann, J., Schumann, G., Borschet, G., Gosseringer, R., Bach, M., Bertling, W.M., Marschalek, R. & Dingermann, T. (1991) Transfer RNA genes from *Dictyostelium discoideum* are frequently associated with repetitive elements and contain consensus boxes in their 5′-flanking and 3′-flanking regions. *J. Mol. Biol.* **222**, 537–552.

26. Marschalek, R., Hofmann, J., Schumann, G., Gösseringer, R. & Dingermann, T. (1992) Structure of DRE, a retrotransposable element which integrates with position specificity upstream of *Dictyostelium discoideum* tRNA genes. *Mol. Cell. Biol.* **12**, 229–239.

27. Schumann, G., Zündorf, I., Hofmann, J., Marschalek, R. & Dingermann, T. (1994) Internally located and oppositely oriented polymerase II promoters direct convergent transcription of a LINE-like retroelement, the *Dictyostelium* repetitive element, from *Dictyostelium discoideum*. *Mol. Cell. Biol.* **14**, 3074–3084.

28. Geier, A., Horn, J., Dingermann, T. & Winckler, T. (1996) Nuclear protein factor binds specifically to the 3′-regulatory module of the long-interspersed-nuclear-element-like *Dictyostelium* repetitive element. *Eur. J. Biochem.* **241**, 70–76.

29. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.

30. Adessi, C., Chapel, A., Vinçon, M., Rabilloud, T., Klein, G., Satre, M. & Garin, J. (1995) Identification of major proteins associated with *Dictyostelium discoideum* endocytic vesicles. *J. Cell Sci.* **108**, 3331–3337.

31. Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal. Biochem.* **72**, 248–254.

32. Wemmer, D.E. & Dervan, P.B. (1997) Targeting the minor groove of DNA. *Curr. Opin. Struct. Biol.* **7**, 355–361.

33. Sastry, M., Fiala, R. & Patel, D.J. (1995) Solution structure of mithramycin–DNA dimers bound to partially overlapping sites on DNA. *J. Mol. Biol.* **251**, 674–689.

34. Gao, X., Mirau, P. & Patel, D.J. (1992) Structure refinement of the chromomycin dimer–DNA oligomer complex in solution. *J. Mol. Biol.* **223**, 259–279.

35. Proudfoot, N. (1991) Poly(A) signals. *Cell* **64**, 671–674.

36. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.

37. Sikorski, R.S., Boguski, M.S., Goebl, M. & Hieter, P. (1990) A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. *Cell* **60**, 307–317.

38. Eckner, R. & Birnstiel, M.L. (1989) Cloning of the cDNAs coding for human HMG-I and HMG-Y proteins: both are capable of binding to the octamer sequence motif. *Nucleic Acids Res.* **17**, 5947–5959.

39. Laurent, B.C., Treitel, M.A. & Carlson, M. (1992) Functional interpedendence of the yeast SNF2, SNF5, and SNF6 proteins in transcriptional activation. *Proc. Natl Acad. Sci. USA* **88**, 2687–2691.