Zinc Dependent Recognition of a Human CpG Island Sequence by the Mammalian Spermatidal Protein TP2[†]

Tapas Kumar Kundu[‡] and Manchanahalli R. Satyanarayana Rao*[,‡,§,||]

Department of Biochemistry, Center for Genetic Engineering, and Jawaharlal Nehru Center for Advanced Scientific Research, Indian Institute of Science, Bangalore 560 012, India

Received May 29, 1996; Revised Manuscript Received September 26, 1996[⊗]

ABSTRACT: Rat spermatidal protein TP2 is a zinc metalloprotein with two atoms of zinc coordinated to cysteine and histidine residues and condenses alternating GC copolymer preferentially in a zinc dependent manner [Kundu, T. K., & Rao, M. R. S. (1995) *Biochemistry 34*, 5143−5150]. In the present study, we have used a 40-mer oligonucleotide containing a human CpG island sequence to study its interaction with TP2 by gel mobility shift assays. A specific complex was observed in the presence of poly(dI)·poly(dC). Preincubation of TP2 with 10 mM EDTA or 1 mM 1, 10-*o*-phenanthroline inhibited the complex formation by more than 90%. Competition experiments with various polynucleotides revealed the following order of efficiency: poly(dG-dC)·poly(dG-dC) > cold homologous oligonucleotide > poly(dA-dT)·poly(dA-dT). Homoduplexes poly(dG)·poly(dC) and poly(dA)·poly(dT) had no effect on the complex formation. Chromomycin A_3, a GC minor groove binding drug, inhibited the complex formation. Methylation of the CpG doublet within the CpG island sequence by *Sss*I methylase (CpG methylase) completely abolished the complex formation. Methylation of G at the N-7 position with dimethyl sulfate did not affect the recognition of CpG island by TP2. Thus, CpG islands, widely distributed in the mammalian genome, may serve as specific loci for initiation of chromatin condensation by TP2 during the later stages of spermiogenesis.

Spermiogenesis in mammals is rather unique in that a set of transition proteins, namely TP1, TP2, and TP4, appear during stages 12−15 (Meistrich, 1989; Unni & Meistrich, 1992). This event is in contrast to direct replacement of the nucleosomal histones by protamines in several other species (Oliva & Dixon, 1991). These transition proteins replace the nucleosomal histones, package the DNA into transcriptionally inert units, and are themselves replaced by the protamine S1, which is the only basic protein present in the final mature epididymal spermatozoa. Elucidation of the processes involved in the displacement of histones by the transition proteins and subsequent packaging is a challenging task by itself, while a more important question is the physiological significance of the evolution of these genes in mammals and the relevance of their appearance during the elongating phase of spermiogenesis. We have been arguing that the function of TP1, TP2, and TP4 may be related to one of the following three important nuclear events that are coincidental with the time of their appearance. They are (1) transformation of nucleosomal-type chromatin into smooth chromatin fiber, (2) initiation of chromatin condensation, and (3) cessation of transcription.

Over the last few years, we have been studying the DNA-binding properties rat TP1 and TP2. Earlier, we had shown that TP1 can melt DNA which is manifested by its two tyrosine residues flanked by basic amino acids (Singh & Rao, 1987). TP1 can also bring about destabilization of nucleosome core particles, suggesting that TP1 may be involved in the displacement of histones from nucleosome-type chromatin (Singh & Rao, 1988). Akama et al. (1995), from their studies on the DNA binding properties of boar TP4, have suggested that TP4 also induces local destabilization of DNA. More recently, we have concentrated our efforts in understanding the DNA binding properties of TP2. To briefly summarize our findings, rat TP2 can condense DNA in contrast to TP1 (Baskaran & Rao, 1990). It is a zinc metalloprotein and contains two atoms of zinc per molecule (Baskaran & Rao, 1991). Both cysteine and histidine residues present in the N-terminal two-thirds of the protein are involved in coordination with zinc (Kundu & Rao, 1994). In a more recent study, we have shown that TP2 in its zinc-coordinated form condenses alternating GC-containing synthetic polynucleotide much more effectively than other types of polynucleotides (Kundu & Rao, 1995). On the basis of this observation, we asked the question to ourselves whether alternating GC-containing sequences are the *in vivo* target sequence for TP2 to initiate condensation within the mammalian genome. One of the well-characterized alternating GC-containing sequences in the mammalian genome is the CpG island. These CpG islands are associated with 5′ or 3′ domains of all known housekeeping genes and also a set of tissue specific genes (Antequera & Bird, 1993). Therefore, we set out to examine the interaction of rat TP2 with an oligonucleotide containing a human CpG island sequence by gel mobility shift assays. The results presented here have

[†] This work was supported by the Department of Atomic Energy. T. K. K. was a Senior Research Fellow of the University Grants Commission.

* Corresponding author: Prof. M. R. S. Rao, Department of Biochemistry, Indian Institute of Science, Bangalore 560 012, India. Phone: 91-80-3092547. Fax: 91-80-3341683. E-mail: mrsrao@biochem.iisc.ernet.in.

[‡] Department of Biochemistry.

[§] Center for Genetic Engineering.

[||] Jawaharlal Nehru Center for Advanced Scientific Research.

[⊗] Abstract published in *Advance ACS Abstracts,* November 1, 1996.

S0006-2960(96)01271-8 CCC: $12.00

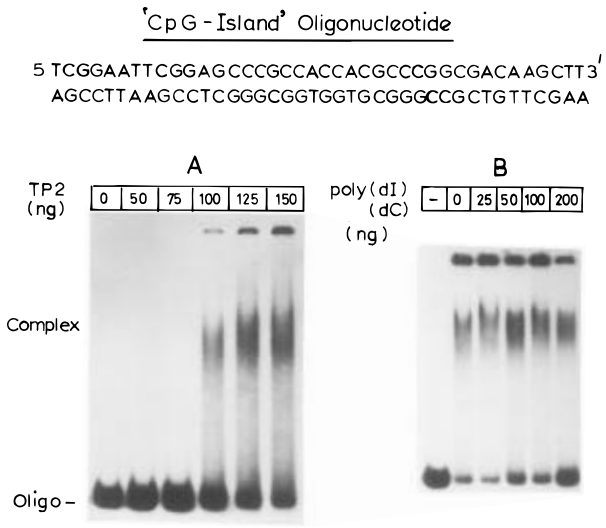

FIGURE 1: Gel retardation assay (GRA) with rat TP2 and the human CpG island oligonucleotide. The sequence in the upper part of the figure is a 40-mer oligonucleotide sequence defined as a CpG island that is present in the downstream of the human ALA dehydratase gene used for GRA. The binding reaction was done with ^{32}P-end-labeled oligonucleotide (35 000 cpm, 2 ng) as described in the text. After the reaction, the complexes were analyzed on an 8% nondenaturing polyacrylamide gel: (A) GRA with increasing concentrations of TP2 and (B) GRA with 100 ng of TP2 in the presence of increasing concentrations of poly(dI)·poly(dC) as a nonspecific competitor.

shown that TP2 recognizes this sequence in a zinc dependent manner. Furthermore, methylation of the oligonucleotide at the C residue of the CpG doublet by CpG (*Sss*I) methylase abolished the specific recognition.

MATERIALS AND METHODS

TP2 was purified from the testes of 60-day-old rats (Wistar) by affinity chromatography on an organomercurial-Sepharose column as described (Kundu & Rao, 1994, 1995). The synthetic polynucleotides—poly(dG-dC)·poly(dG-dC), poly(dA-dT)·poly(dA-dT), poly(dG)·poly(dC), poly(dA)·poly(dT), poly(dI-dC)·poly(dI-dC), and poly(dI)·poly(dC)—were purchased from Pharmacia Biotech Limited. Restriction enzymes and other DNA-modifying enzymes were from either New England Biolabs or Bangalore Genei (Bangalore, India). CpG methylase (*Sss*I methylase) was from New England Biolabs. The complementary oligonucleotides (40-mer) containing the human CpG island sequence 3′ downstream of the ALA dehydratase gene were synthesized on a Pharmacia gene assembler. They were purified on a 6 M urea-containing 18% polyacrylamide gel. [γ-^{32}P]ATP (specific activity of >3000 Ci/mmol) was from Bhabha Atomic Research Center (Bombay, India).

Gel Mobility Shift Assays. The assay was carried out according to the method of Ye and Samuels (1987) with several modifications. The sequence of the oligonucleotide containing the human CpG island sequence used in our binding studies is shown in Figure 1. The upper strand was labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase and annealed with an equimolar concentration of the complementary strand in 10 mM Tris-HCl (pH 8.0)/1 mM EDTA/20 mM NaCl by heating at 95 °C for 15 min and slowly cooling at room temperature over several hours. The duplex oligonucleotide was purified on a native 60% PAGE. The binding reaction was carried out in a 20 μL volume containing 10 mM Tris-HCl (pH 7.5)/5 mM $MgCl_2$/50 mM KCl/12.5% glycerol/0.1% Triton X-100. In each reaction, 100 ng of TP2 incubated with 10 μM $ZnSO_4$ and 35 000 cpm of the oligonucleotide (approximately 2 ng) was used. The labeled probe was added last, and the binding reaction was carried out at 25 °C for 30 min. After the reaction, 4 μL of 5 × dye (0.05% bromophenol blue and 50% glycerol) was added and the samples were electrophoresed on an 8% native polyacrylamide gel (acrylamide:bisacrylamide = 29:1) in 6.6 mM Tris-HCl (pH 7.5) and 3.3 mM sodium acetate (pH 7.5). The gel was preelectrophoresed for 30 min at 10 V/cm or 20 mA current at 15 °C. After the samples were loaded, electrophoresis was carried out for 2 h. The gel was subsequently dried and autoradiographed. For competition experiments, the competitor nucleic acid was added after the addition of TP2 and incubated at 25 °C for 10 min. The labeled oligonucleotide was added in the end. In the chemical interference assays, the drug chromomycin A_3 was added soon after adding the protein and before adding the labeled probe.

To examine the effect of methylation of the G residue at the N-7 position and the C residue at the C-5 position within the CpG doublet, on the recognition of the CpG island sequence by TP2, the duplex oligonucleotide was methylated with dimethyl sulfate (DMS) and CpG methylase (*Sss*I methylase), respectively. The G residue methylation by DMS was performed according to Wissmann and Hillen (1991), and the reaction was monitored by piperidine cleavage and urea−PAGE. Methylation of the cytosine residue was carried out with CpG methylase according to the manufacturer's (New England Biolabs) recommendations. Briefly, 1 μg of CpG island oligonucleotide was incubated with 5 units of *Sss*I methylase in a reaction buffer containing 50 mM NaCl, 10 mM Tris-HCl (pH 7.9), 10 mM $MgCl_2$, 1 mM DTT, and 160 μM *S*-adenosylmethionine for 60 min at 37 °C. The oligonucleotide was purified by phenol extraction and ethanol precipitation.

RESULTS

TP2 Binds the CpG Island Sequence in a Zinc Dependent Manner. On the basis of our recent observation that the rat spermatidal protein TP2 condenses alternating GC copolymer much more efficiently than other types of nucleic acids (Kundu & Rao, 1995), we set out to examine whether TP2 also recognizes a natural GC rich sequence, namely the CpG island. For this purpose, we selected a sequence 3′ downstream of the ALA dehydratase (porphobilinogen synthetase) gene for our study. An oligonucleotide (40-mer) corresponding to this sequence, which contains six CpG doublets, was synthesized and used in all our experiments. Panels A and B of Figure 1 show the results of the gel mobility shift experiments carried out with this 40-mer oligonucleotide and TP2. It can be seen from Figure 1A that the complex started appearing at 100 ng of TP2. The amount of the complex formed increased with input TP2 concentration of up to 150 ng. Above 150 ng, the complexes formed were in the form of aggregates which were retained in the well. Even at 125 and 150 ng of TP2, a small percentage of the complex was retained in the well. Figure 1B shows the results obtained with increasing concentrations of poly(dI)·poly(dC) as a nonspecific nucleic acid competitor keeping the TP2 concentration constant at 100 ng. It is evident that up to 100

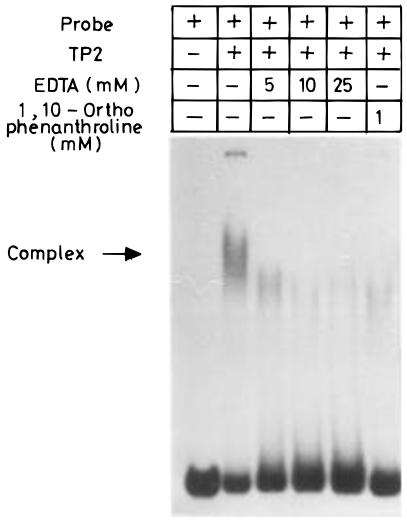

FIGURE 2: Effect of removal of zinc from TP2 on its binding to the CpG island oligonucleotide. TP2 was preincubated with either EDTA (5, 10, and 25 mM) or 1,10-*o*-phenanthroline (1 mM) and then used in the binding reaction.

ng of poly(dI)·poly(dC) there was no inhibition of the complex formed, indicating that the complexes are specific in nature. There was, however, some competition observed at 200 ng of poly(dI)·poly(dC). A small increase in the free probe observed with 50 ng of the competitor might be due to variation in gel loading of the samples. The optimum KCl concentration required for the complex formation was found to be 50 mM, beyond which there was a smear in the retarded band (data not shown). The results presented in Figure 1A also indicated that there is a narrow window of TP2 to DNA ratio wherein one can observe the complexes in the form of a retarded band, beyond which the complexes are pulled toward formation of aggregates. We tried several incubation and gel electrophoretic conditions to prevent the formation of aggregates without any success, indicating that it is an inherent property of TP2−DNA complexes. This inherent property of TP2−DNA complex to aggregate at higher TP2 concentrations, therefore, did not permit us to calculate the association constant by this method.

Since TP2 is a zinc metalloprotein (Baskaran & Rao, 1991), we next examined whether the recognition of the CpG island oligonucleotide by TP2 is zinc dependent. TP2 was preincubated with different concentrations of EDTA (5, 10, and 25 mM) or with 1,10-*o*-phenanthroline (1 mM), a specific chelator of zinc, and then used for binding experiments. Both these metal chelators reduced the formation of the protein−oligonucleotide complex significantly (Figure 2). A densitometric scan of this autoradiogram revealed that EDTA at 5, 10, and 25 mM decreased the amount of complex formed by 80, 90, and 95%, respectively. There was a 90% decrease in the amount of complex formed with *o*-phenanthraline. It can also be noted that even the residual complex observed in EDTA- or 1,10-*o*-phenanthroline-treated samples (lanes 3−6) moved ahead of that observed with untreated TP2 (lane 2). The reason for this behavior is not clear, and it is quite likely that the complexes formed in the absence of zinc are not tight like the one observed in lane 2. Alternately, it is also possible that the complexes observed may be a mixture of several bands, among which one of them is more resistant to EDTA treatment. However, these results definitely suggest that zinc is essential for TP2 to recognize the CpG island oligonucleotide. We tried to reconstitute the EDTA-treated TP2 with zinc to restore the binding to the CpG island oligonucleotide. However, these experiments were unsuccessful, possibly due to inherent problems in handling small amounts of the protein and inefficient removal of EDTA by dialysis. We have now been able to express rat TP2 in *Escherichia coli* (Meetei & Rao, 1996), and we are now in the process of studying in great detail the role of zinc in the recognition of CpG island oligonucleotide by TP2.

Effect of Synthetic Polynucleotides of Defined Sequence as Competitors for CpG Island Oligonucleotide Recognition by TP2. The sequence preference exhibited regarding DNA recognition by TP2 was established by adding different synthetic polynucleotides as competitors in the gel mobility shift experiments. We had shown earlier that TP2 condenses alternating copolymer poly(dG-dC)·poly(dG-dC) 3-fold more efficiently than it condenses poly(dA-dT)·poly(dA-dT), while the homoduplexes poly(dG)·poly(dC) and poly(dA)·poly(dT) were condensed to a much lesser extent (Kundu & Rao, 1995). We used these polynucleotides in our competition studies. The results of these experiments are shown in Figure 3A−F. As expected, cold homologous oligonucleotide was effective as a competitor (Figure 3A). Interestingly, the alternating copolymer poly(dG-dC)·poly(dG-dC) was a better competitor than even the cold homologous oligonucleotide (Figure 3B). This can be explained by the fact that this alternating copolymer provides many more GC binding sites for TP2 than the CpG island oligonucleotide. Figure 3C shows the results obtained with the alternating copolymer poly(dA-dT)·poly(dA-dT) as the competitor. Although there was an initial decrease in the amount of complex formed (0.7-fold excess), it reached a plateau with higher concentrations of the polynucleotide. When the two homoduplexes poly(dG)·poly(dC) and poly(dA)·poly(dT) were used as competitors in the gel mobility shift assays, we found that both these homoduplexes could not compete with the complex formation (Figure 3D,E). We have also used in our competition experiments poly(dI-dC)·poly(dI-dC) to see whether the G residue in poly(dG-dC)·poly(dG-dC) is important in the recognition process of the CpG island sequence by TP2. As can be seen from Figure 3F, this alternating copolymer was not effective as a competitor when compared to poly(dG-dC)·poly(dG-dC). We point out here that poly(dI-dC)·poly(dI-dC) exists as a closely related B type of DNA structure (Wang & Keildering, 1993). Therefore, it is possible that this altered helical nature of poly(dI-dC)·poly(dI-dC) is responsible for the lack of competition.

A quantitative analysis of the experiments observed with alternating copolymers is shown in Figure 4. The autoradiograms of panels A−C and F of Figure 3 were scanned, and the percentage of the complex was plotted as a function of fold excess of the competitor. The competition curve with poly(dA-dT)·poly(dA−dT) shows a biphasic pattern. The initial phase of competition probably represents inhibition of recognition by TP2 of an alternating purine-pyrimidine stretch, while the plateau region of the curve represents a more specific recognition of the alternating GC stretch which cannot be competed out by poly(dA-dT)·poly(dA-dT). It is interesting to note that the binding characteristics of TP2 with the CpG island oligonucleotide as revealed by these competition experiments with various synthetic polynucle-

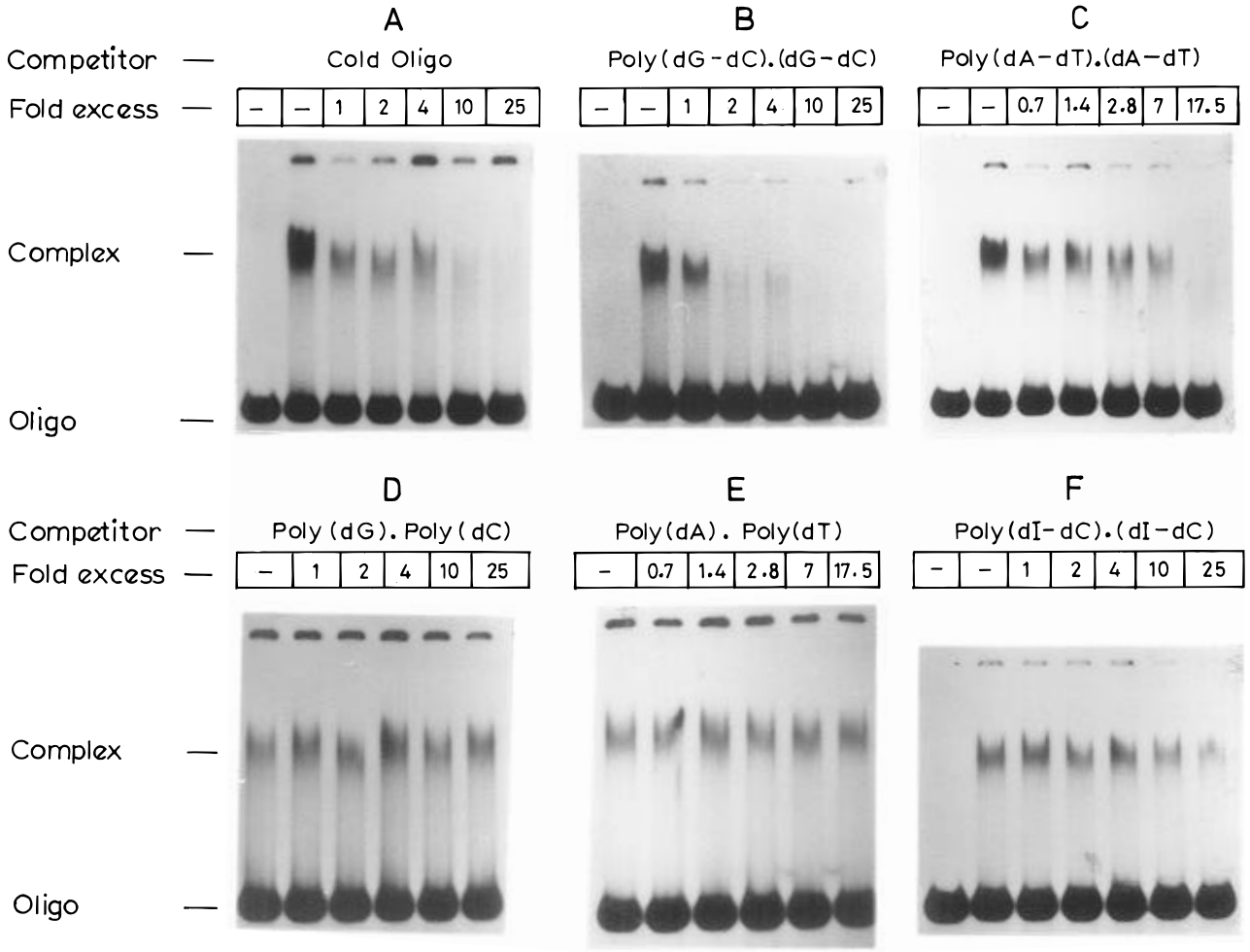

FIGURE 3: Effect of synthetic polynucleotides of defined sequences on the binding of TP2 to the human CpG island oligonucleotide. GRA was done as described in the legend to Figure 1 in the presence of increasing concentrations of (A) cold oligonucleotide, (B) poly(dG-dC)·poly(dG-dC), (C) poly(dA-dT)·poly(dA-dT), (D) poly(dG)·poly(dC), (E) poly(dA)·poly(dT), and (F) poly(dI-dC)·poly(dI-dC). The fold molar excess indicated is with respect to nucleotide concentration.

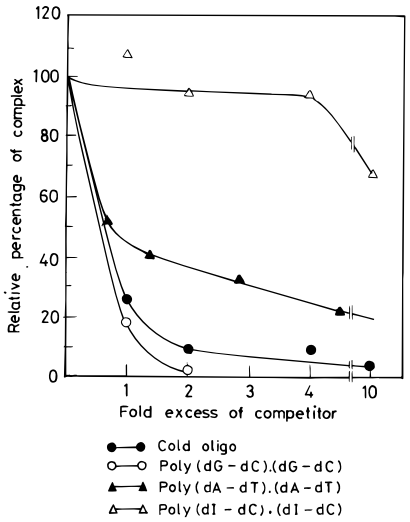

FIGURE 4: Quantitative analysis of competetion of the TP2−CpG island oligonucleotide complex with different synthetic alternating copolymers. The autoradiograms of Figure 3 were scanned in an LKB laser densitometer. The value obtained for the control lane (without any competitor) was taken as 100%. The relative percentage decrease in complex formation was calculated and plotted. Each value is an average of two independent experiments.

otides parallel the observations we had made earlier while studying the condensation behavior using circular dichroism spectroscopy (Kundu & Rao, 1995). While the CD spectroscopic studies monitored changes in the secondary and tertiary structure of DNA in the nucleoprotein complexes, the results presented here directly examine the primary recognition process leading to the formation of the nucleoprotein complex.

TP2 Binds to the Minor Groove of the CpG Island Sequence. Chemical interference in the gel mobility shift assay has been extensively used to determine the groove specificity of DNA binding proteins (Dickinson et al., 1992). Since the experiments described above suggested that TP2 binds preferentially GC rich sequences, we have used chromomycin A_3 which binds specifically to the GC minor groove (Banville et al., 1990; Keniry et al., 1993). In this experiment, we incubated the CpG island oligonucleotide with different concentrations of chromomycin A_3 before adding TP2 to the binding reaction mixture. Addition of the drug (20 μM) to the binding reaction mixture in the absence of TP2 resulted in the formation of a drug−DNA complex which migrated a little more slowly than the free oligonucleotide probe (Figure 5, lane 7). Chromomycin A_3 at 0.5 and 1 μM did not appreciably affect the formation of TP2−CpG island complex (lanes 5 and 6). However, at 5 and 20 μM, the drug completely inhibited complex formation (lanes 3 and 4) by blocking the minor groove in the CpG island oligonucleotide as shown by the formation of the

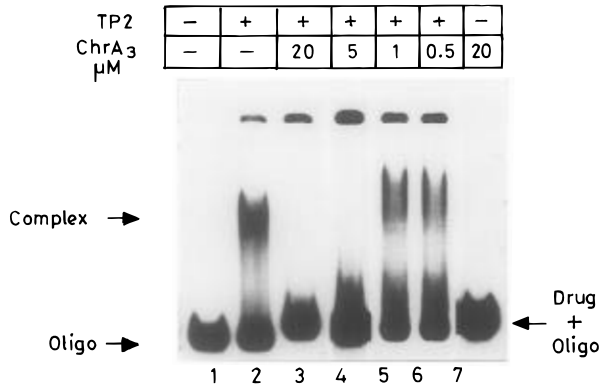

FIGURE 5: Effect of chromomycin A_3 on TP2 binding to the human CpG island oligonucleotide. Chromomycin A_3 (Chr A_3) was added at the start of the binding reaction at the micromolar concentration indicated above the lanes. The TP2−DNA and chromomycin A_3 (drug)−DNA complexes are indicated by the arrows. The binding reaction was done as described in the text and legend to Figure 1 in the presence of 200 ng of poly(dI)·poly(dC) as a nonspecific competitor.

drug−DNA complex. These results, therefore, suggest that TP2 binds to the GC minor groove within the CpG island sequence.

Effect of Methylation of C and G Residues in the CpG Island Sequence on Its Recognition by TP2. The CpG islands are methylation free zones in the mammalian genome (Antequera & Bird, 1993). Methylation of CpG islands inhibits the binding of CpG island binding proteins and controls gene expression (Boyes & Bird, 1993). In order to determine the effect of methylation of the human CpG island upon recognition by TP2, the C residues within the CpG dinucleotides were methylated at the C-5 position by using CpG methylase (*Sss*I methylase). The CpG island oligonucleotide contains one CCGG sequence whose methylation was monitored by its resistance to *Hpa*II digestion (Figure 6A). Methylation of the oligonucleotide completely inhibited formation of the TP2−CpG island oligonucleotide complex (Figure 6B). Despite the common belief that the influence of the methyl groups is through direct methyl group protein interactions, it is not necessary that the only means by which methyl groups modulate protein−DNA interactions is through direct contact. On the basis of their investigation on the effect of N-6 methylation of A on the binding of *Eco*RI, Diekmann and Mclaughlin (1988) have suggested that methyl groups may regulate the binding in part by altering the conformation of DNA within the domain that is interacting with the proteins. Recently, Hodges-Garcia and Hagerman (1992) have shown that cytosine methylation can induce local distortion in the structure of duplex DNA. Keeping this possibility in mind, we examined whether methylation of the CpG island oligonucleotide by *Sss*I methylase altered the minor groove dimensions by carrying out binding studies with chromomycin A_3. The results shown in Figure 6C reveal that the methylated CpG island oligonucleotide did bind chromomycin A_3 with the same efficiency as the nonmethylated oligonucleotide, indicating that the groove dimensions are not significantly altered upon methylation of the C residue.

The N-7 position of the G residues in the CpG island oligonucleotide was methylated with dimethyl sulfate. Methylation was ascertained by checking the piperidine cleavage product on a polyacrylamide sequencing gel (data not

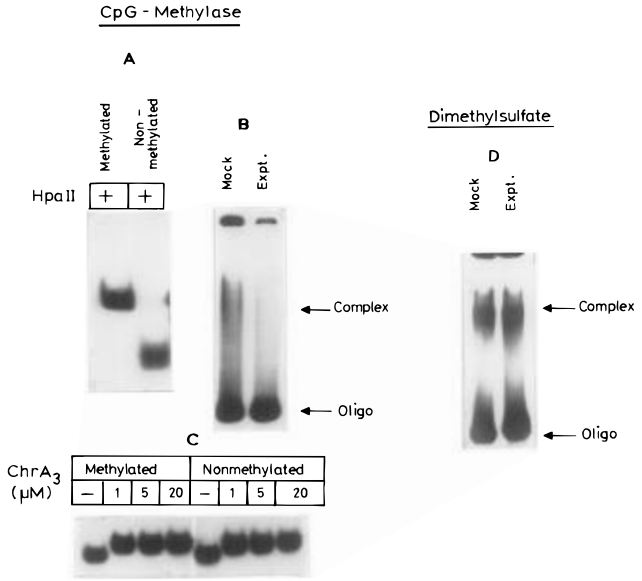

FIGURE 6: Effect of C-5 methylation of cytosine in the CpG doublet and N-7 methylation of guanine residue on the recognition of the CpG island oligonucleotide by TP2. (A) Monitoring the methylation of the C-5 position of cytosine. CpG island oligonucleotide was methylated by *Sss*I methylase (CpG methylase) as described by the NEB protocol. ^{32}P-end labeled methylated and mock-treated oligonucleotides were digested with restriction endonuclease *Hpa*II and analyzed on an 8% nondenaturing polyacrylamide gel. (B) Autoradiogram of a gel retardation assay for TP2 complex formation with mock-treated and methylated CpG island oligonucleotide. The binding reaction was carried out as described in the legend to Figure 1 in the presence of 200 ng of poly(dI)·poly(dC). (C) Autoradiogram of a gel retardation assay of chromomycin A_3 binding to mock-treated and CpG-methylated oligonucleotide. The micromolar concentrations of chromomycin A_3 used are indicated at the top of the lanes. (D) Autoradiogram of a gel retardation assay for TP2 complex formation with mock-treated or DMS-treated CpG island oligonucleotide. The binding reaction was done as described above.

shown). In contrast to the effect of methylation of the cytosine residue, N-7 methylation of the G residue did not affect the formation of TP2−CpG island oligonucleotide complex (Figure 6D).

DISCUSSION

In this paper, we have studied the interaction of rat spermatidal protein TP2 with a human CpG island sequence to test the hypothesis that they may represent the *in vivo* target sequence for initiation of chromatin condensation during mammalian spermiogenesis. We could not use a rat CpG island sequence as they are not well characterized. However, CpG island sequences are reported to be more or less well conserved within the mammalian species (Antequera & Bird, 1993). The results of the experiments described here clearly show that rat TP2 forms a specific complex with the CpG island oligonucleotide. The sequence preferential binding of TP2 was established by competition experiments using several synthetic polynucleotides as competitor. The observation that poly(dG-dC)·poly(dG-dC) was more effective as a competitor than even the homologous cold CpG island oligonucleotide strengthens our conclusion that TP2 does prefer alternating GC stretches for binding. Since the homoduplex poly(dG)·poly(dC) did not compete at all, one can logically interpret that GC base pair within the CpG doublet, and not a single GC base pair is important

for recognition of the CpG island oligonucleotide by TP2.

An important observation made in this study is the requirement of zinc in the recognition of the CpG island oligonucleotide by TP2. Our earlier studies had shown that TP2 has predominantly type I β turn structure contributed by several proline residues present in the protein (Kundu & Rao, 1994). Removal of zinc from the protein resulted in alteration in the CD spectrum, indicating that zinc induces/stabilizes a particular conformation in the protein. The present results provide direct evidence for the involvement of zinc-induced/stabilized polypeptide fold of TP2 in recognition of DNA.

The TP2−CpG island oligonucleotide complex was observed only over a narrow window of the TP2 to DNA ratio. Formation of the aggregates at a higher TP2/DNA ratio is an inherent property of TP2 due to the presence of several lysine and arginine residues present in the C-terminal one-third of the molecule. We have suggested earlier that TP2 may be considered to have two domains (Kundu & Rao, 1995). The N-terminal domain having the zinc binding residues may serve as the DNA recognition domain, while the C-terminal domain may serve as the DNA condensation domain. We are now in the process of studying recognition of the CpG island sequence by the N-terminal domain alone so that we can understand the kinetic parameters of the binding process which was not possible with intact TP2.

The present investigation has also given some insights into the possible mode of interaction of TP2 with GC rich DNA. A chemical interference assay with chromomycin A_3 has clearly indicated that one of the modes of interaction is through the minor groove of GC rich DNA. It is generally believed that sequence specific DNA binding proteins interact with DNA in the major groove through their α helices, while nonsequence specific DNA binding proteins like histones bind through the minor groove (Travers, 1993). However, there are exceptions to this general rule (Raumann et al., 1994). Recent studies on the cocrystal and NMR structures of several zinc finger motifs and their recognition sequences have shown the importance of both the major and minor grooves in the recognition process (Schmeideskamp & Kleit, 1994). Taking together our results on methylation and chemical interference experiments, we can suggest that TP2 may interact with the minor groove while also making a contact with the C residue which is projected toward the major groove. A more refined structural study of the zinc-induced/stabilized polypeptide fold in TP2 by using two-dimensional NMR and X-ray crystallographic techniques is necessary for a better understanding of the interaction of TP2 with the CpG doublet in the CpG islands. However, these physical techniques require large amounts of the protein for analysis. Toward this goal, we have recently been successful in expressing native rat TP2 in *E. coli* (Meetei & Rao, 1996). A fine resolution map of the interaction of TP2 with the CpG island sequence at the nucleotide level is also being studied with the N-terminal (1−86 amino acids) fragment of TP2 to circumvent the problem of aggregation encountered with intact TP2.

In the mammalian genome, CpG islands are associated with almost all the housekeeping genes present in the 5′-flanking promoter elements and 3′ downstream domains (Antequera & Bird, 1992). Our observation that the mammalian spermatidal protein TP2 which appears during stages 10−15 of spermiogenesis binds to a CpG island sequence in a zinc dependent manner may have important implications with events that occur at the time of its appearance. Firstly, on the basis of its DNA-condensing property, TP2 may initiate condensation using the CpG island as the target sequence. Secondly, since CpG islands are present in the promoter domains of the housekeeping genes, binding of TP2 and ensuing condensation may also result in repression of transcriptional activity, an event that is also coincidental with the time of appearance of the transition proteins.

The appearance of transition proteins TP1, TP2, and TP4 during spermiogenesis is unique to mammals. In most other species, there is a direct transition from nucleosome-type chromatin to nucleoprotamine fiber (Oliva & Dixon, 1991). Several *in vivo* studies have shown that protamines from rooster, squid cuttlefish, and fish can directly replace the histones from chromatin (Oliva et al., 1987). Why only mammals have evolved an intermediate stage during spermiogenesis involving transition proteins has long been a puzzle. Recently, the genomic organization of human (Nelson & Krawetz, 1993) and rat (Adham et al., 1991) TP2 has been described. Human TP2 is a part of the gene cluster housing protamine P1, protamine P2, and TP2, in that order (Nelson & Krawetz, 1993). The P1 class of protamines is widely distributed, while P2 protamine and TP2 are unique to mammals. Because of tight linkage of these genes, it has been suggested that P2 and TP2 genes might have arisen by gene duplication of the P1 protamine gene. However, since there is no extensive similarity between the P2 and TP2 genes, it has been speculated that they might have undergone accelerated evolution following a gene duplication event. In this connection, it is also interesting to note that the CpG island is also highly evolved in mammals (Antequera & Bird, 1993). Thus, it is tempting to speculate that there might be a close correlation between the evolution of CpG islands and TP2 in mammals.

ACKNOWLEDGMENT

We thank Dr. S. K. Brahmachari, Convenor of Oligonucleotide Facility at IISc, for the synthesis of CpG island oligonucleotides.

REFERENCES

Adham, J. M., Szpirer, C., Kremling, H., Keime, S., Szpirer, J., Leran, G., & Engel, W. (1991) *Cytogenet. Cell Genet. 57*, 47−51.

Akama, K., Ichimura, H., Sato, H., Kojima, S., Miura, K., Hayashi, H., Komatsu, Y., & Nakano, M. (1995) *Eur. J. Biochem. 233*, 179−185.

Antequera, F., & Bird, A. (1993) In *The Chromosome* (Hoslop-Harrison, J. S., & Flavell, R. B., Eds.) pp 127−132, Bios Scientific Publishers, Oxford, U.K.

Banville, D. L., Keniry, M. A., Kam, M., & Shafer, R. H. (1990) *Biochemistry 29*, 6521−6534.

Baskaran, R., & Rao, M. R. S. (1990) *J. Biol. Chem. 265*, 21039−21047.

Baskaran, R., & Rao, M. R. S. (1991) *Biochem. Biophys. Res. Commun. 179*, 1491−1499.

Boyes, J., & Bird, A. (1992) *EMBO J. 11*, 327−333.

Dickinson, L. A., Joh, T., Kohwi, Y., & Kohwi-Shigematsu, T. (1992) *Cell 70*, 631−645.

Diekman, S., & McLaughlin, L. W. (1988) *J. Mol. Biol. 202*, 823−834.

Hodges-Garcia, Y., & Hagerman, P. J. (1992) *Biochemistry 31*, 7595−7599.

Keniry, M. A., Banville, D. L., Simonds, P. M., & Shafer, R. (1993) *J. Mol. Biol. 231*, 753−767.

Kundu, T. K., & Rao, M. R. S. (1994) *FEBS Lett. 351*, 6−10.
Kundu, T. K., & Rao, M. R. S. (1995) *Biochemistry 34*, 5143−5150.
Meetei, A. R., & Rao, M. R. S. (1996) *Protein Expression Purif.* (in press).
Meistrich, M. L. (1989) in *Histones and other basic Nuclear proteins* (Hnilica, L. S., Stein, G. S., & Stein, J. L., Eds.) pp 165−182, CRC Press, Boca Raton, FL.
Nelson, J. E., & Krawetz, A. (1993) *J. Biol. Chem. 268*, 2932−2936.
Oliva, R., & Dixon, G. H. (1991) *Prog. Nucleic Acid Res. Mol. Biol. 40*, 25−94.
Oliva, R., Bazett-Jones, D., Mezquita, C., & Dixon, G. H. (1987) *J. Biol. Chem. 262*, 17016−17025.
Raumann, B. F., Brown, B. M., & Sauer, R. T. (1994) *Curr. Opin. Struct. Biol. 4*, 36−43.
Schmiedeskamp, M., & Kleuit, R. E. (1994) *Curr. Opin. Struct. Biol. 4*, 28−35.
Singh, J., & Rao, M. R. S. (1987) *J. Biol. Chem. 263*, 734−740.
Singh, J., & Rao, M. R. S. (1988) *Biochem. Int. 17*, 701−710.
Travers, A. A. (1993) in *DNA-Protein interactions*, Chapman and Hall, London.
Unni, E., & Meistrich, M. L. (1992) *J. Biol. Chem. 267*, 25359−25363.
Wang, L., & Keiderling, T. A. (1993) *Nucleic Acids Res. 21*, 4127−4132.
Wissmann, A., & Hillen, W. (1991) *Methods Enzymol. 208*, 365−375.
Ye, Z. S., & Samuels, H. H. (1987) *J. Biol. Chem. 262*, 6313−6317.

BI961271I