THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 275, No. 49, Issue of December 8, pp. 38500–38507, 2000

*Printed in U.S.A.*

# Identification of Two Novel Zinc Finger Modules and Nuclear Localization Signal in Rat Spermatidal Protein TP2 by Site-directed Mutagenesis*

Received for publication, March 31, 2000, and in revised form, August 24, 2000
Published, JBC Papers in Press, August 28, 2000, DOI 10.1074.jbc.M002734200

**Amom Ruhikanta Meetei‡, Kolathur S. Ullas, and Manchanahalli R. Satyanarayana Rao§**

*From the Department of Biochemistry, Indian Institute of Science, Bangalore 560 012, India*

**Spermatidal protein TP2, which appears transiently during stages 12–16 of mammalian spermiogenesis, is a DNA condensing zinc metalloprotein with a preference to GC-rich DNA. We have carried out a detailed site-directed mutagenesis analysis of rat spermatidal protein TP2 to delineate the amino acid residues involved in coordination with two atoms of zinc. Two zinc fingers modules have been identified involving 4 histidine and 4 cysteine residues, respectively. The modular structure of the two zinc fingers identified in TP2 define a new class of zinc finger proteins that do not fall into any of the known classes of zinc fingers. Transfection experiments with COS-7 cells using wild type and the two zinc finger pocket mutants have shown that TP2 preferentially localizes to nucleolus. The nuclear localization signal in TP2 was identified to be $^{87}$GKVSKRKAV$^{95}$ present in the C-terminal third of TP2 as a part of an extended NoLS sequence.**

Spermiogenesis is an elaborate process of cellular differentiation wherein the haploid spermatids produced following meiotic division finally mature into spermatozoa. In mammals, in particular, it is characterized by the transient appearance of a group of proteins, namely TP1, TP2, and TP4 which replace the somatic and testis-specific histones, before themselves being replaced by protamines (1). The biological significance of the role of transition proteins in chromatin remodeling, in mammals, is a matter of great importance to understand the final stages of sperm development and maturation in the testis. Among the three basic proteins, TP1, TP2, and TP4, TP1 was shown to be a DNA melting protein *in vitro* and has been speculated to be involved in the destabilization of nucleosome to facilitate displacement of histones (2, 3). More recently Yu *et al.* (4) have shown that TP1-deficient mice have abnormal spermatogenesis with reduced fertility. TP4 has been shown recently to stimulate SV40 DNA relaxing activity of eukaryotic DNA topoisomerase I suggesting a probable role in chromatin reorganization (5). On the other hand, more information is available regarding DNA binding properties of TP2. TP2, in contrast with TP1, was shown to have DNA condensing properties (6). Rat TP2 is a zinc metalloprotein, containing two atoms of zinc per molecule (7), condenses GC-rich DNA more than other types of DNA sequences (8), and recognizes a human CpG island sequence in a zinc-dependent manner (9). Methylation of cysteine residue in the CpG doublet inhibited the recognition of the CpG island sequence by TP2. Chromomycin $A_3$ interference experiments revealed that probably one of the modes of interaction of TP2 with GC-rich DNA is through its minor groove. These results suggested that CpG islands might serve as specific loci for initiation of chromatin condensation by TP2 through its zinc-binding domain.

For a better understanding of the nature of interaction of TP2 with DNA, it is necessary to delineate the zinc-binding domains of TP2. A careful examination of the amino acid sequence of rat TP2 and its alignment with the known zinc fingers did not reveal any similarity. In addition, TP2 has many more number of histidine and cysteine residues than that are necessary for coordinating two atoms of zinc. We, therefore, undertook the challenge to identify the amino acid residues involved in zinc coordination by employing extensive site-directed mutagenesis of cysteine and histidine residues present in TP2. The results presented here reveal novel zinc finger modules in TP2. We have further carried out transient transfection experiments in COS-7 cells with wild type and mutant forms of TP2 and the results show that wild type TP2 preferentially localizes to nucleolus. We have also identified the amino acid stretch that is necessary for nuclear localization (NLS)[1] of TP2.

## EXPERIMENTAL PROCEDURES

*Expression of TP2 and Its Mutants in Escherichia coli, Purification of Recombinant Proteins*—The cDNA cloning of TP2, generation of partial synthetic cDNA, and optimization of its expression in *E. coli* were reported recently (10, 11). Since TP2 has cysteine and histidine residues in its N-terminal two-thirds portion (Fig. 1), several single and multiple mutants of TP2 have been generated wherein histidine and cysteine residues were changed to glutamine and alanine, respectively. All the mutants of TP2 carrying multiple mutations were generated in a single PCR product as described previously (12). The desired mutations were confirmed by DNA sequencing in ABI 377 automated DNA sequencer using dye terminator chemistry. The wild type and mutants of TP2 were purified from isopropyl-1-thio-β-D-galactopyranoside-induced BL21(DE3) cells, in a single step using heparin-agarose column instead of a combination of $Zn^{2+}$ affinity and heparin-agarose chromatography (11). Briefly, the cells were lysed in a buffer containing 50 mM Tris-HCl, pH 8.0, 0.57 M NaCl, 10 mM β-mercaptoethanol, and 0.2 mM phenylmethylsulfonyl fluoride. The supernatant collected after centrifugation at 120,000 × *g* for 1 h was loaded onto the heparin-agarose column, which was pre-equilibrated with the lysis buffer. After washing the column with 10 column volumes of the lysis buffer, the bound protein

* This work was supported in part by the Department of Biotechnology, Government of India. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Senior Research Fellow of the Council of Scientific and Industrial Research. Present address: Laboratory of Genetics, NIA, National Institutes of Health, Baltimore, MD 21224-6820.

§ To whom correspondence should be addressed: Dept. of Biochemistry, Indian Institute of Science, Bangalore 560 012, India. Tel.: 91-80-3092547; Fax: 91-80-3600814 or 3600118; E-mail: mrsrao@biochem.iisc.ernet.in.

[1] The abbreviations used are: NLS, nuclear localization signal; PCR, polymerase chain reaction; GST, glutathione *S*-transferase; PAGE, polyacrylamide gel electrophoresis.

was eluted with 3 column volumes of the same buffer containing 1 M NaCl. The eluted protein was desalted and concentrated.

*Construction of GST-TP2 Fusion Vectors and Purification of GST-TP2 Peptides*—To check the ability of the N-terminal 1–43-amino acid region of TP2 to bind zinc, the cDNAs spanning the corresponding region of wild type and mutant TP2s were PCR amplified and subcloned into *Bam*HI and *Sma*I sites of pGEX-2T expression vector (Amersham Pharmacia Biotech). The expressed GST-TP2 fusion proteins were purified through GSH-Sepharose column. Similarly, another GST-TP2 protein spanning the 40–86-amino acid region was also generated. Purification of GST fusion proteins were carried out by standard methods using GSH-Sepharose 4B (Amersham Pharmacia Biotech) column.

*Quantitative Radioactive Zinc ($^{65}ZnCl_2$) Blotting of Wild Type and TP2 Mutants*—Both wild type and mutant TP2s (5 $\mu$g each) were blotted onto a nitrocellulose membrane strip which was pre-soaked in buffer A (10 mM Tris-HCl, pH 7.5, 300 mM NaCl, 10 mM EDTA, and 2 mM DTT). The membrane strip was incubated at 37 °C for 1 h and washed with binding buffer B (10 mM Tris HCl, pH 7.5 and 300 mM NaCl) three times (15 min each time) to remove excess EDTA and dithiothreitol. The membrane was subsequently incubated in buffer B containing 30 $\mu$Ci of $^{65}ZnCl_2$ (specific activity, 800 mCi/g, Bhabha Atomic Research Center, Mumbai) at room temperature for 30 min with gentle shaking. The unbound radioactivity was removed by washing the membrane three times with buffer B (20 min each time), dried, and subjected to autoradiography. The radioactive spots were scanned and quantitated using Bio-Rad Gel DOC 1000 system. The intensity of the spot thus obtained was corrected for the amount of the protein as calculated from densitometric scanning of the same amounts of protein run on a SDS-polyacrylamide gel (13). The background signal obtained without any protein was subtracted from the experimental values.

*Cell Culture, Transfection, and Immunofluorescence*—COS-7 cells (African Green Monkey Kidney cells) were maintained in Dulbecco's modified Eagle's medium/F-12 (1:1) containing 10% fetal bovine serum at 37 °C with 5% $CO_2$. The wild type and mutant cDNAs of TP2 were cloned into mammalian expression vector pCMX PL-1 for transfection studies. The sequence $^{87}$GKVAKRKAV$^{95}$ has homology to the consensus NLS (14). In order to confirm that this sequence harbors NLS, a C-terminal deletion mutant having amino acid residues 1–95 was generated using sequence-specific primers by PCR and then cloned into pCMX PL-1. A mutant lacking the NLS-(87–95) was also generated to support the same. The N-terminal fragment 1–86 and the C-terminal fragment 96–114 were separately generated by PCR using specific primers and then cloned into pCMX PL-1. COS-7 cells were grown on coverslips in 24-well plates and the different constructs were transfected using LipofectAMINE PLUS reagent (Life Technologies, Inc.). Immunofluorescence was carried out 12 h after transfection. Briefly, the cells were fixed with 4% paraformaldehyde in phosphate-buffered saline for 20 min at room temperature followed by permeabilization using 0.1% Triton X-100 for 10 min. The cells were then treated with 1% normal goat serum in phosphate-buffered saline for 45 min at room temperature followed by incubation with monospecific rabbit anti-rat TP2 antibodies for 45 min. The cells were then incubated with secondary antibody (fluorescein isothiocyanate-conjugated goat anti-rabbit IgG). The cells were counter-stained with 4,6-diamidino-2-phenylindole and visualized in a Zeiss Axioplan microscope.

## RESULTS

The cDNA derived amino acid sequence of rat TP2 is shown in Fig. 1. The protein has 5 cysteine and 11 histidine residues most of which are distributed in the N-terminal two-thirds of the molecule. There is a convenient glutamate residue at the 86$^{th}$ position. Cleavage of TP2 with V8 protease generates the N-terminal two-thirds and C-terminal one-third and we have conveniently used this division in all our experimental approach. We developed a single step purification procedure for the purification of recombinant wild type and mutant TP2 (partial synthetic cDNA) expressed in *E. coli* based on the very basic nature of the C-terminal region of TP2. The SDS-PAGE pattern of the different fractions is shown in Fig. 2*A*. The bound TP2 protein that could be eluted from heparin-agarose column with 1 M NaCl showed a single band on the polyacrylamide gel.

*Identification of Zinc Binding Pockets in TP2*—The single and multiple mutants of TP2 involving cysteine and histidine residues were generated by the method described recently (12). The BL21(DE3) cells harboring the various mutants were induced with

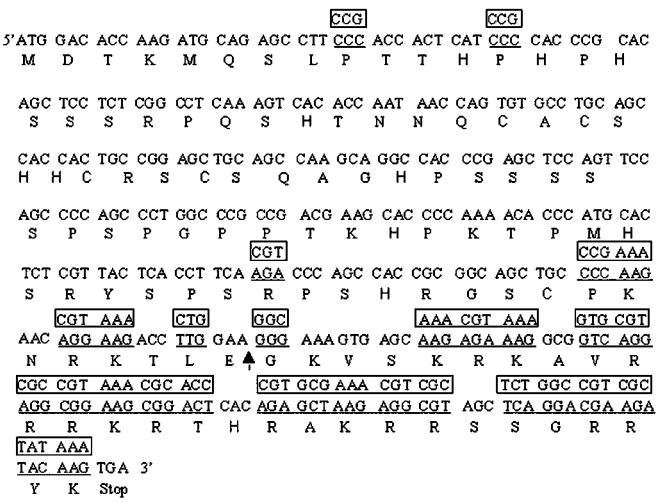


FIG. 1. **Rat TP2 cDNA and its derived amino acid sequence.** Histidine and cysteine residues are shown in *bold* and the *arrow* indicates the single glutamate residue (V8 Protease cleavage site) which has been used arbitrarily to define the N-terminal two-thirds and C-terminal one-third of TP2. The sequences shown in *boxes* are the codons, which were incorporated to generate partial synthetic TP2 cDNA for optimal expression in *E. coli* (10).

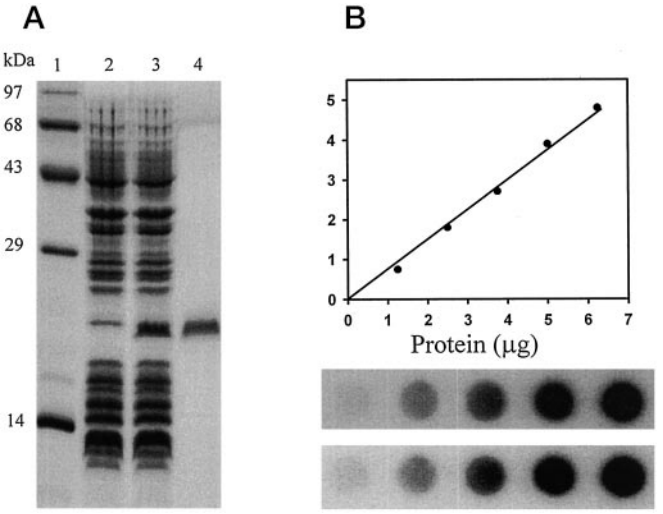


FIG. 2. **Purification of recombinant TP2 and quantitative $^{65}$zinc blotting.** *A,* single step purification of TP2 by heparin-agarose chromatography. *Lane 1*, molecular weight marker; *lane 2*, uninduced *E. coli* extract; *lane 3*, induced *E. coli* extract; and *lane 4*, heparin-agarose eluate. *B,* quantitative zinc blotting. Wild type TP2 (0–6 $\mu$g) was spotted on nitrocellulose membrane in duplicates and processed for $^{65}$zinc binding as described under "Experimental Procedures." The radioactive spots were quantified using Bio-Rad Gel Documentation 1000 system. The average of the two independent blots was plotted as a function of TP2 concentration (*upper panel*).

isopropyl-1-thio-$\beta$-D-galactopyranoside and purified as described above. Since the N-terminal two-thirds of TP2-(1–86) can still bind zinc with equal efficiency as native TP2 (15), the histidine residue at position 102 was not included in the mutagenesis studies. The wild type and mutant TP2 proteins were checked for their zinc binding capacity by quantitative $^{65}$zinc dot blot analysis. The quantitative nature of zinc binding to TP2 was established by monitoring the linearity of $^{65}$zinc binding to TP2 as a function of increasing concentration of TP2. A duplicate analysis of such a $^{65}$zinc blot is shown in the *bottom panel* of Fig. 2*B* while the linearity curve as analyzed by densitometric scan is shown in the *upper panel* of Fig. 2*B*. It is clear that the method employed is linear in the range of protein concentration used in these experiments. Radioactive zinc blotting has also been used recently to analyze the zinc binding

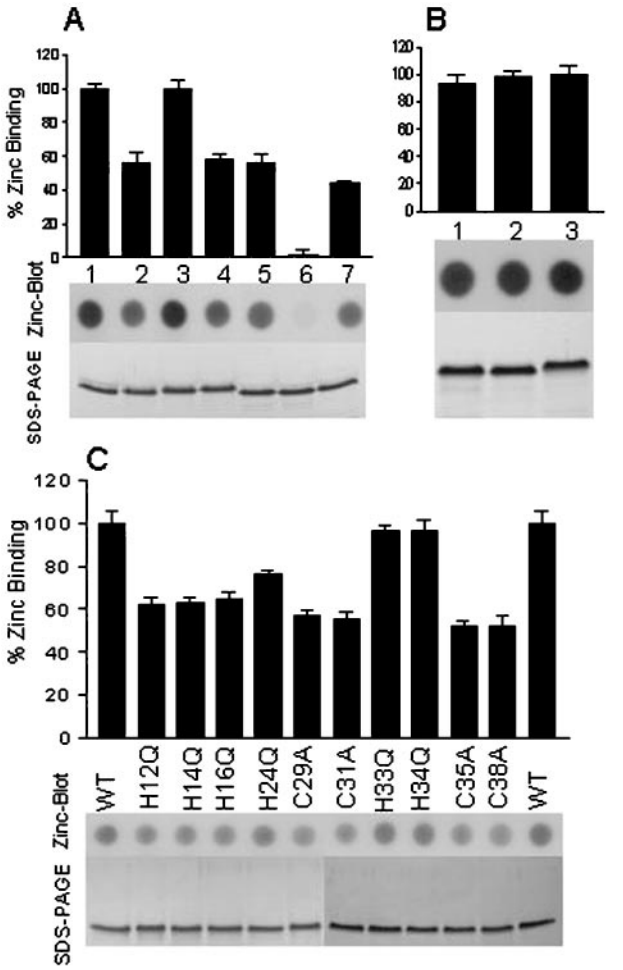


FIG. 3. **Analysis of $^{65}$zinc binding capacity of the wild type and mutant TP2.** *A-C,* $^{65}$zinc blot analysis of multiple mutants of TP2. The zinc binding activity of the wild type TP2 was plotted as 100% against mutants. In each of the panels, Coomassie Blue-stained protein in an SDS-PAGE is also given to show the purity as well as the quantity of TP2 used for zinc-blot analysis. *A, lane 1*, wild type TP2; *lane 2*, H12Q,H16Q,H24Q mutant; *lane 3*, H43Q,H58Q,H64Q,H74Q mutant; *lane 4*, H12Q,H16Q,H24Q,H33Q,H34Q,H43Q,H58Q,H64Q,H74Q mutant; *lane 5*, C29A,C31A,C35A,C38A mutant; *lane 6*, H12Q,H16Q, H24Q/C29A,C31A,C35A,C38A mutant; *lane 7*, C29A,C31A,C35A,C38A/ H43Q,H58Q,H64Q,H74Q mutant. *B, lane 1*, H58Q,H64Q/C78A mutant; *lane 2*, H43Q,H58Q,H64Q,H74Q mutant; and *lane 3*, wild type TP2. *C,* $^{65}$zinc blot analysis of single site mutations of TP2. The nature of the mutant TP2 is given *below* the bar diagram. The values represented in the *bar* diagram is an average ± S.D. of three independent binding experiments. The background signal obtained without any protein was subtracted from all the experimental values.

efficiencies of RNA polymerase II subunits and their mutants (16) using very similar procedure. In all the subsequent experiments, TP2 (wild type and mutants) were spotted at 5 $\mu$g concentration and the purity of the proteins as analyzed by SDS-PAGE is shown in each of the figures. Fig. 3 shows the $^{65}$zinc binding analysis of wild type TP2 and its various mutants. The *middle panel* shows the autoradiogram obtained after $^{65}$zinc binding. The histogram showing the extent of $^{65}$zinc binding by the various mutants in comparison to wild type TP2 (taken as 100%) is shown at the *top* of each figure. Each of the binding experiments was done in triplicate and the percent binding shown in the figure is an average ± S.D of these independent experiments. The observation made in Fig. 3*A* can be summarized as follows. The H43Q,H58Q,H64Q,H74Q mutant did not affect zinc binding. On the other hand, H12Q,H16Q,H24Q, and H12Q,H16Q,H24Q,H33Q,H34Q,H43Q, H58Q,H64Q,H74Q mutants had only 56 and 58% zinc binding capacity, respectively, as compared with wild type TP2. Since TP2 binds two atoms of zinc, a reduction of 40–50% $^{65}$zinc binding can be taken as the loss of one zinc-binding site. Thus, these results suggest that His$^{12}$, His$^{16}$, and His$^{24}$ are involved in binding one atom of zinc while there was no contribution from His$^{43}$, His$^{58}$, His$^{64}$, and His$^{74}$. The fourth histidine residue involved in the coordination of zinc in this module was identified as His$^{14}$ by single mutant studies in the background of other zinc pocket mutant (see below), as multiple mutants could not be generated involving His$^{14}$ due to the overlapping sequence in the mutagenic primers used in this region. Among the cysteine mutants, the C29A,C31A, C35A,C38A mutant had only 56% zinc binding capacity (*lane 5*) indicating that it has lost one zinc-binding site retaining the site involving histidine residues. The involvement of each of the two zinc binding pockets was further confirmed by the observation that the double pocket mutant, H12Q,H16Q,H24Q/C29A,C31A,C35A, C38A mutant lost 98.5% of the zinc binding capacity (*lane 6*). The non-involvement of His$^{43}$, His$^{58}$, His$^{64}$, and His$^{74}$ in zinc coordination is evident from *lane 7* wherein mutation at these positions did not further decrease the zinc binding activity to near 0%. As mentioned above, there are 5 cysteine residues in TP2. The results described above indicate that Cys$^{29}$, Cys$^{31}$, Cys$^{35}$, and Cys$^{38}$ form the second pocket for coordination with zinc. To rule out any contribution of the fifth cysteine (Cys$^{78}$) residue, mutants were created to include Cys$^{78}$ and the results obtained with H58Q,H64Q/C78A showed only about 6% reduction in zinc binding (Fig. 3*B, lane 1*).

In the next series of experiments, single mutants of TP2 were created to substantiate the conclusions drawn above. The histidine mutants at positions 12, 14, and 16 had approximately 62, 63, and 64% zinc binding activity, respectively, compared with wild type TP2 while H33Q,H34Q did not have any effect on zinc binding (Fig. 3*C*). The H24Q mutant had approximately 76% zinc binding activity. Theoretically one would have expected a near 50% reduction in zinc binding activity. It is possible that the fourth coordinating molecule in the H24Q mutant might be partially served by a water molecule (17). Among the cysteine mutants, C29A, C31A, C35A, and C38A resulted in a 43, 48, 48, and 48% reduction in zinc binding activity, respectively, clearly indicating that they are involved in zinc coordination in TP2. These results further indicate that all the four residues in each of the zinc-binding pocket is necessary for zinc coordination. Even a loss of a single coordination site abolishes the zinc binding activity of that pocket.

Fig. 4, *A* and *B,* show the results of experiments wherein mutations were created at each of the coordinating amino acids of one pocket while the other pocket was a complete mutant. It is clear from Fig. 4*A* that H12Q,H16Q,H24Q (*lane 2*), H12Q,H16Q,H24Q,H33Q (*lane 5*), and H12Q,H16Q,H24Q, H34Q (*lane 6*) had lost 40–50% zinc binding activity. With this mutant in background (H12Q,H16Q,H24Q), a single mutation each at Cys$^{29}$, Cys$^{31}$, Cys$^{35}$, and Cys$^{38}$ resulted in a complete loss of zinc binding activity (*lanes 3, 4, 7,* and *8*). The fact that each of the secondary mutants at the cysteine residue resulted in a complete loss of zinc binding activity, it can be concluded that these cysteine residues form one zinc binding pocket. A similar set of experiments was done for single histidine mutants in the background of the C29A,C31A,C35A,C38A/ H43Q,H58Q,H64Q,H74Q mutant. It was shown in Fig. 3*B* that His$^{43}$, His$^{58}$, His$^{64}$, and His$^{74}$ are not involved in zinc coordination. The results presented in Fig. 4*B* show that mutations at His$^{12}$, His$^{14}$, and His$^{16}$ resulted in a total loss of zinc binding activity. The zinc binding activity of H24Q (*lane 6*) was 28% although this also should have resulted in a total loss of zinc binding activity. As discussed above, a water molecule may be serving for the fourth coordination ligand for zinc.

To provide further evidence for the two zinc binding pockets in TP2, the first 43 amino acids of TP2 were expressed as a GST

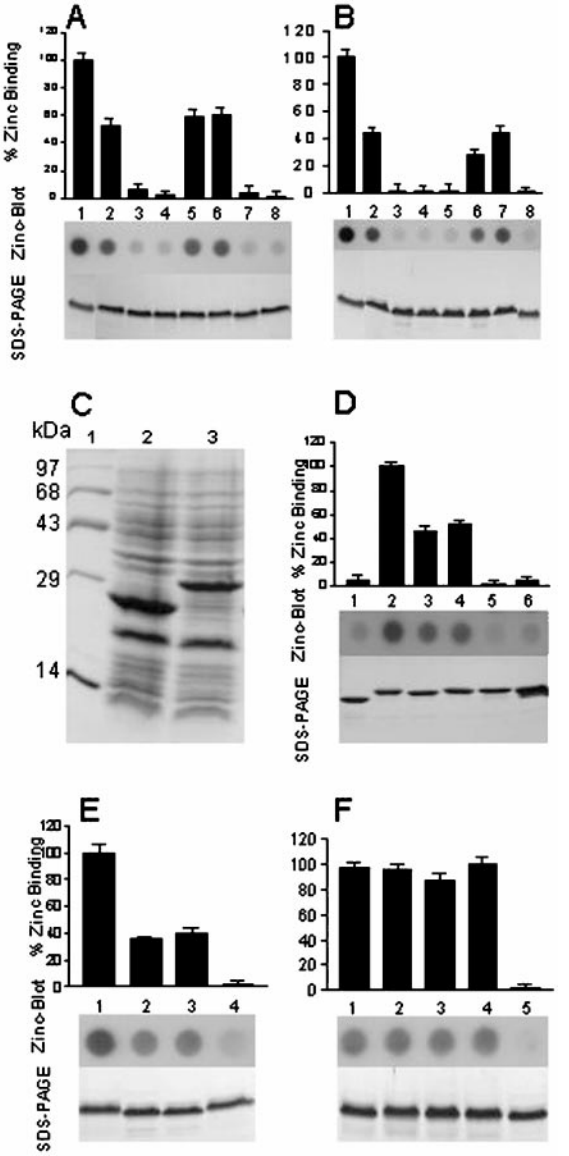


FIG. 4. **$^{65}$Zinc blot analysis of different mutants of TP2.** *A,* cysteine to alanine mutations in the background of H12Q,H16Q,H24Q mutant. *Lane 1*, wild type TP2; *lane 2*, H12Q,H16Q,H24Q mutant; *lane 3*, H12Q,H16Q,H24Q/C29A mutant; *lane 4*, H12Q,H16Q,H24Q/C31A mutant; *lane 5*, H12Q,H16Q,H24Q,H33Q mutant; *lane 6*, H12Q, H24Q,H34Q mutant; *lane 7*, H12Q,H16Q,H24Q/C35A mutant; and *lane 8*, H12Q,H16Q,H24Q/C38A mutant. *B,* histidine to glutamine mutations in the background of C29A,C31A,C35A,C38A/H43Q,H58Q,H64Q, H74Q (X) mutant. *Lane 1*, wild type TP2; *lane 2*, X mutant; *lane 3*, H12Q-X mutant; *lane 4*, H14Q-X mutant; *lane 5*, H16Q-X mutant; *lane 6*, H24Q-X mutant; *lane 7*, H33,H34Q-X mutant; and *lane 8*, H12Q,H16Q,H24Q-X mutant. *C,* induction profile of GST-partial TP2 fusion protein. *Lane 1*, molecular weight marker; *lane 2*, pGEX-2T induced *E. coli* extract; and *lane 3*, pGEX-2T TP2 induced *E. coli* extract. *D,* the $^{65}$zinc blot with the GST partial TP2 fusion protein and its mutants was carried out as described in the legend to Fig. 2. *Lane 1*, GST; *lane 2*, GST wild type partial TP2 (1–43 residues); *lane 3*, GST partial TP2 containing the cysteine pocket mutation; *lane 4*, GST partial TP2 containing the histidine pocket mutation; *lane 5*, GST partial TP2 containing both cysteine and histidine pocket mutations; and *lane 6*, GST partial TP2 (43–86 residues). *E,* proline to alanine mutants in the background of cysteine pocket mutant. *Lane 1*, wild type TP2; *lane 2*, P13A,P15A/C29A,C31A,C35A,C38A mutant; *lane 3*, C29A,C31A, C35A,C38A mutant; and *lane 4*, histidine and cysteine double pocket mutant. *F,* arginine to alanine mutant in the background of wild type TP2. *Lane 1*, R20A mutant; *lane 2*, R36A mutant; *lane 3*, R20A,R36A mutant; *lane 4*, wild type TP2; and *lane 5*, histidine and cysteine double pocket mutant. The values represented in the *bar* diagram are an average ± S.D. of three independent binding experiments. The background signal obtained without any protein was subtracted from all the experimental values.

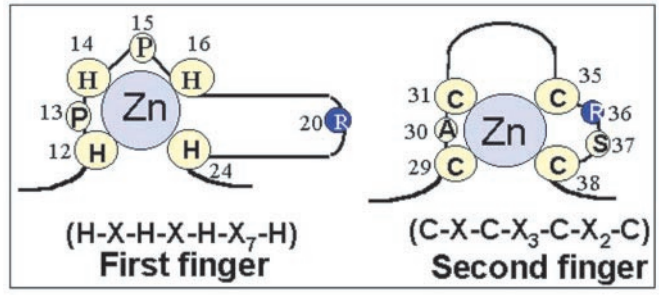


FIG. 5. **Zinc finger modules of TP2.**

fusion protein and analyzed for their zinc binding capacity. Fig. 4*C* shows the SDS-PAGE pattern of *E. coli* cells harboring the expression plasmid with and without isopropyl-1-thio-β-D-galactopyranoside induction. A 30-kDa band was clearly seen in the induced lane corresponding to the GST-TP2 fusion protein which was subsequently purified by affinity chromatography. The *bottom panel* in Fig. 4*D* shows the SDS-PAGE of the wild type fusion protein and the various mutants derived from it. The GST alone did not bind any $^{65}$zinc, however, the GST-TP2 fusion protein did bind $^{65}$zinc. Both the cysteine and histidine mutant peptides lost about 54 and 48% $^{65}$zinc binding activity, respectively. The total histidine and cysteine mutant lost all the $^{65}$zinc binding activity. A GST fusion protein, which had TP2 sequence (43–86), did not possess any $^{65}$zinc binding activity (*lane 6*).

Based on all these results, we propose the zinc finger modules coordinating the two zinc atoms in TP2 as shown in Fig. 5. The TP2 zinc finger modules that we have delineated do not correspond to any of the known zinc finger modules present in steroid and thyroid hormone receptors or RING finger or transcription factors TFIIIA and GATA 1 or binuclear zinc cluster of GAL 4. The major features of the zinc finger modules of TP2 are the following: 1) the first finger is made up of four coordinating histidine residues, 2) there is only one amino acid between the first and the second and the second and the third coordinating histidine residues that happens to be proline. Its mutation to alanine, however, did not affect zinc binding activity (Fig. 4*E*). There are seven amino acids between the third and the fourth coordinating histidine residues out of which there is a lone arginine residue at position 20. In many of the known zinc fingers, it has been demonstrated that zinc fingers make contact with nucleic acid bases (mostly guanine) predominantly through an arginine residue (18, 19). 3) The second zinc finger module is made up of four cysteine residues. 4) There is a single amino acid between the first and the second coordinating cysteine residues. 5) Three amino acids separate the second and the third coordinating cysteine residues. 6) Two amino acids separate the third and the fourth coordinating cysteine residues and one happens to be arginine at position 36. Mutation of either of the arginine residues 20 or 36 to alanine did not affect zinc binding activity (Fig. 4*F*). 7) There are no conserved amino acids such as Phe, Leu, Ile, and Val in both the modules as observed in other zinc finger proteins.

*Transient Expression of Wild Type and Mutant TP2 in COS-7 Cells*—To understand the *in vivo* physiological significance of TP2 and its various DNA-binding domains, we have begun to carry out ectopic expression analysis of TP2 and its mutants in COS-7 cells. Since TP2 is expressed only in the male germ cells during spermiogenesis and the fact that the process of spermatogenesis involves several waves of germ cells differentiating asynchronously, the effect of its ectopic expression in cells in culture should give valuable information on this protein. An ideal situation of studying the effect of expression of TP2 would

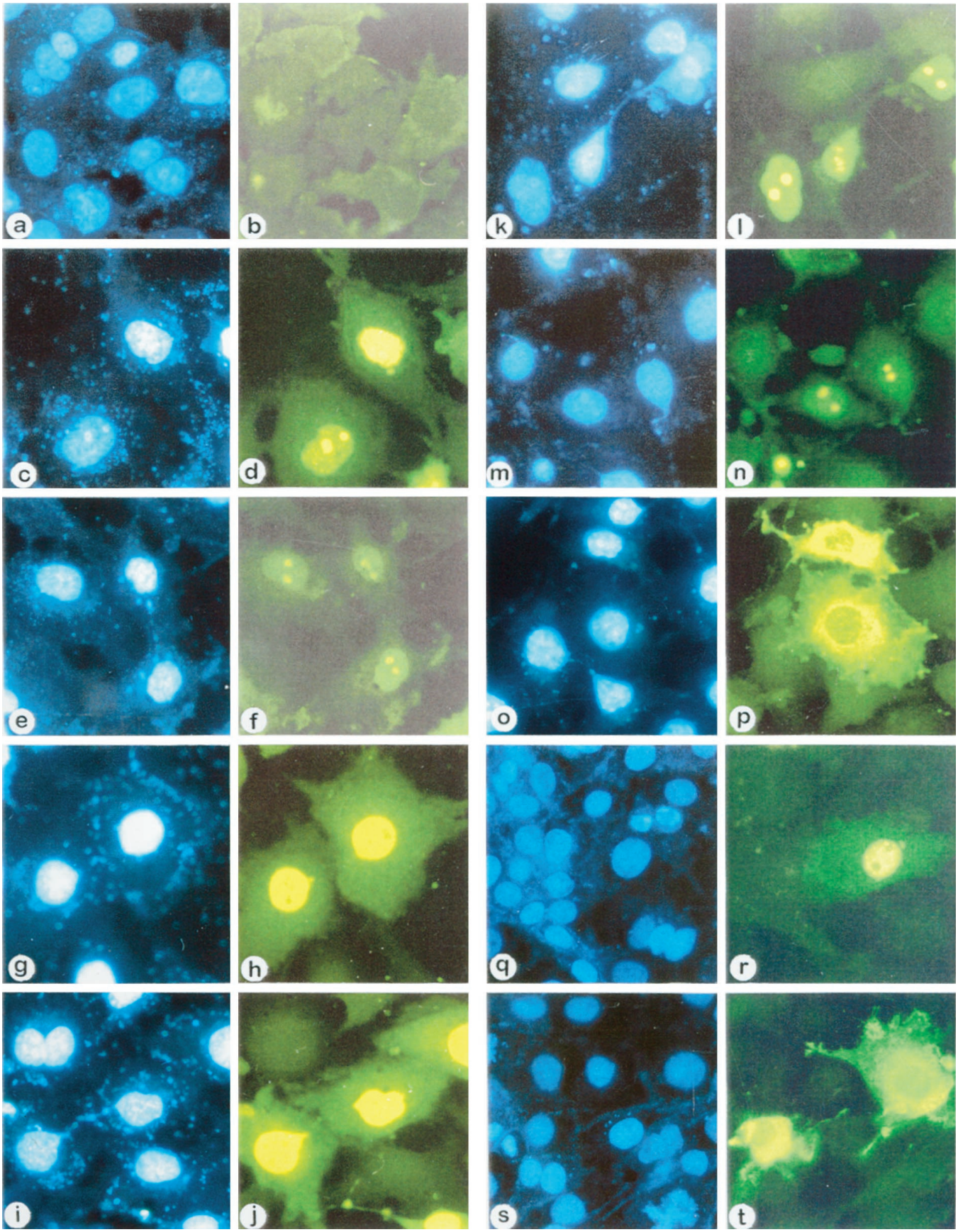


FIG. 6. **Immunofluorescence of COS-7 cells transfected with wild type TP2 and its mutants.** Wild type and mutant rat TP2 cDNAs, in pCMX PL1, were transfected into COS-7 cells as described under "Experimental Procedures." 12 h after transfection, immunofluorescence was carried out using monospecific rabbit anti-rat TP2 antibody. *Panels d, f, h, j, l, n, p, r,* and *t* show the localization pattern of wild type TP2 and its mutants, Wild type TP2 (*d*), histidine pocket mutant (*f*), cysteine pocket mutant (*h*), histidine and cysteine double pocket mutant (*j*), R20A mutant (*l*), R36A mutant (*n*) C-terminal deletion mutant, (*p*) C-terminal deletion mutant with NLS-(1–86) (*r*) and TP2 without the NLS (87–95 deletion

be with round spermatids. However, culturing of round spermatids isolated from testicular cells, *in vitro*, has not been demonstrated yet. In the absence of such a culture system, we have used COS-7 cells as the experimental model system. Although TP2 is unique to rat testis and not expressed in a somatic cell, such a study would definitely give some insights into the effect of expression of TP2 on cellular events in general. For this purpose, we subcloned the wild type TP2 (containing the natural codons) and its various mutants into pCMX PL1 whose expression is driven by the constitutive cytomegalovirus promoter. The expression of wild type TP2 and its mutants in COS-7 cells was monitored by immunofluorescence using monospecific rabbit anti-TP2 antibodies. These antibodies reacted with both wild type and mutant TP2s with equal efficiency in a Western blot analysis (data not shown). The immunofluorescence patterns as well as the corresponding 4,6-diamidino-2-phenylindole-stained nuclei are shown in Fig. 6. The most interesting observation that can be made from this figure is that wild type TP2 preferentially localizes to nucleolus (*panel d*). There are two cells showing the immunofluorescence pattern of TP2 with different levels of expression. When the expression is low, TP2 is predominantly localized to nucleolus. When there is a higher level of expression of TP2, it is also observed in nucleoplasm in addition to the nucleolus. The expression and localization pattern of the zinc finger pocket mutants also reveal some interesting features. There was no appreciable difference in the localization pattern of histidine pocket mutant because the mutant TP2 also preferentially localized in the nucleolus (*panel f*). However, interestingly, the cysteine pocket mutant did not show such preferential localization to nucleolus and most of the expressed protein was observed in the nucleoplasm (*panel h*). The double histidine and cysteine pocket mutants behaved like the cysteine pocket mutant (*panel j*). The localization pattern of the two arginine mutants (R20A and R36A) behaved like wild type TP2 (*panels l* and *n*). As mentioned earlier, there is a single glutamate residue at position 86 based on which we have assigned the N-terminal two-thirds as the DNA recognition domain with its two zinc fingers and the C-terminal one-third having stretches of basic amino acid residues as the DNA condensation domain (8, 9). When a TP2 construct in pCMX PL1 vector containing only the N-terminal two-thirds of the molecule-(1–86) was transfected into COS-7 cells, this partial TP2 was predominantly localized in the cytoplasm (*panel p*) indicating the presence of a potential nuclear localization signal in the C-terminal third of TP2 (Fig. 8). Analysis of the amino acid sequence of TP2 revealed that the amino acid stretch $^{87}$GKVSKRKAV$^{95}$ had strong homology to monopartite NLS (14). We therefore, generated TP2 construct in pCMX PL-1 vector containing amino acid residues 1 to 95. When transfected into COS-7 cells, this fragment was now localized to the nucleus (*panel r*) indicating that this sequence indeed corresponds to NLS in TP2. Interestingly TP2 was found only in the nucleoplasm but not in the nucleolus. To further corroborate this conclusion, we deleted this stretch of amino acids from TP2 and when transfected with this deletion mutant, the TP2 was again found in the cytoplasm (*panel t*) similar to that observed in the C-terminal deletion mutant (1–86).

## DISCUSSION

The present investigation is mainly concerned with the identification of the amino acids that are involved in zinc coordination in TP2. For this purpose we have taken the approach of site-directed mutagenesis to create several mutants of TP2 and then evaluating their $^{65}$zinc binding capacity in a quantitative zinc blot analysis. Since multiple mutants had to be created in TP2 involving several combinations of cysteine and histidine residues, we recently developed a novel technique for generating several multiple site-specific mutants of TP2 without a need for cloning each of the mutant generated (12). The radioactive zinc binding studies reported here have clearly identified the cysteine and histidine residues involved in zinc coordination and the emerging model for the two zinc finger modules is shown in Fig. 5 which do not fit into the known zinc finger modules. The $Cys_2$-$His_2$ fingers constitute one of the most important and versatile families of eukaryotic DNA-binding domains, which are present in numerous transcription factors that recognize a diverse set of DNA sequences (18, 19). These fingers have provided an important model system for studying the principles of protein-DNA recognition and offer a useful framework for the selection and design of novel DNA-binding proteins (20). The crystal structures of the zinc finger domains with their target DNA-binding sites have revealed a generally conserved DNA-locking arrangement with the $\alpha$-helix (recognition helix) fitting into the major groove (21, 22). Amino acids at three key positions on the surface of the recognition helix: $-1$, 3, and 6 (numbering with respect to start of the $\alpha$ helix) play a dominant role in defining sequence specificity making one to one amino acid to base contacts to a base triplet (contacts on the primary strand involve a three base pair subsite).

With this information in background, it is important to discuss the contrasting features of the zinc finger modules of TP2 and its secondary structure. TP2 does not have much $\alpha$-helicity as predicted by various algorithms as well as by circular dichroism spectroscopic experiments (8, 15). It has a higher propensity to form $\beta$-turns due to the presence of large number of prolines. Although the zinc stabilized structure involves predominantly a type I $\beta$-turn structure, a small induced $\alpha$-helical component was not ruled out. Such spectroscopic studies have now revealed similar results with boar TP2 (23, 24). Therefore, the structural features of TP2 along with its novel zinc finger modules are completely different from those of the other zinc finger proteins. However, we would like to point out here a major difference, between TP2 and other zinc finger proteins. TP2 shows only sequence preference binding to alternate GC-rich DNA. On the other hand, other known classical zinc finger proteins are sequence-specific DNA-binding proteins. Hence the molecular recognition and docking of the zinc finger modules of TP2 with the DNA helix may be completely different from those observed with known classical zinc finger proteins. Some of the preliminary data obtained by us earlier do point out some features of TP2 and DNA interaction. For example, chromomycin $A_3$ interference experiments on the TP2 recognition of the CpG island suggests that TP2 might interact with alternating GC-rich containing DNA predominantly through the minor groove (9). However, based on the observation that methylation of cytosine residue in the CpG doublet abolished the binding of TP2 to the CpG island sequence, contribution of the major groove also in the recognition process cannot be ruled out since 5-methyl-cytosine projects toward the major groove in the DNA helix. A more detailed study through NMR/crystallography is necessary to understand the recognition mechanism underlying TP2-DNA interaction.

We have also now begun to ask questions as to the effects of ectopic expression of TP2 in a cell culture system. For this

mutant) (*t*). Negative control, COS-7 cells transfected with vector alone, is shown in *panel b*. *Panels a, c, e, g, i, k, m, o, q,* and *s* show the nuclear staining of the corresponding fields, with 4,6-diamidino-2-phenylindole.

purpose the original rat TP2 cDNA clone (containing the natural codons) was put under the regulation of cytomegalovirus promoter and the expression was followed in COS-7 cells using monospecific TP2 antibodies. As mentioned earlier, in the absence of a suitable *in vitro* cell culture and transfection system available for round spermatids, we have made use of an established cell line, COS-7, for these studies. The most interesting observation of these studies is that wild type TP2 preferentially localizes to nucleolus containing ribosomal DNA, which is GC-rich (25). The fact that the C-terminal deletion mutant is predominantly localized to cytoplasm suggests that a nuclear localization signal is in the C-terminal fragment. Further experiments with different constructs of TP2 have clearly delineated the NLS of TP2 to $^{87}$GKVSKRKAV$^{95}$, which closely resembles the consensus sequence for a monopartite NLS (14). It is generally believed that nucleolar localization signals share some common features namely: 1) the signal is highly rich in basic amino acid residues, especially arginine and 2) the signal is normally linked to the NLS, forming an extended NLS (26, 27). In this respect, TP2 has highly basic amino acid residues (particularly rich in arginine) next to the identified NLS sequence $^{87}$GKVSKRKAV$^{95}$ satisfying the above criteria (Fig. 1). This is also supported by the observation that TP2 lacking the 95–114 amino acid stretches does not localize to nucleolus although it is translocated into the nucleus. Although some reports have suggested that the nucleolar localized proteins also have their specific address signals in their amino acid sequences (28–32), we have been unable to identify such NoLS sequences in TP2. We would like to draw attention here that in addition to the requirement of basic amino acid stretch in the "extended" NLS sequence, even the cysteine pocket mutant and the cysteine-histidine double pocket mutant of TP2 failed to preferentially localize in the nucleolus (Fig. 6, *h* and *j*). Thus, both the cysteine finger module and the extended NLS of TP2 seem to be necessary for nucleolar localization. It is worth mentioning here that there is no physical barrier between the nucleoplasm and nucleolus and it is believed that in most cases the nucleolar localization results from specific protein-protein or protein-nucleic acid interaction. In this context, the requirement of the cysteine finger module for nucleolar localization assumes significance and it is very likely that the targeting may be driven by greater affinity of TP2 to ribosomal DNA which is GC-rich. We believe that this preference *in vivo* to nucleolar rDNA may have significance in the sequence of events taking place during nucleoprotein transition in elongating spermatids. At present, we are developing strategies to culture and do transfection analysis of round spermatids *in vitro* with TP2 cDNA. The present observations, therefore, are giving some initial insights into the events taking place inside the cell nucleus upon TP2 expression which can be interpreted based on the *in vitro* nucleic acid binding properties of TP2 (8, 9). Another intriguing observation that we have made in these studies is that the expression level of the mutant TP2 (both histidine and cysteine mutant) is much more than wild type TP2 (Fig. 6, *h* and *j*). A likely explanation for this observation is that wild type TP2 probably autoregulates its own expression by binding to the cytomegalovirus promoter which contains 32 CpG dinucleotides within a stretch of 610 nucleotides which can qualify for a CpG island. This autoregulation may be relieved in the zinc finger mutants. Such a phenomenon has great significance particularly with the argument we have been proposing that TP2 initiates chromatin condensation during spermiogenesis using CpG island sequences, a consequence of which may be repression of global transcription. Recently, we have also raised an interesting possibility that there may be a close correlation between the evolution of CpG islands and TP2

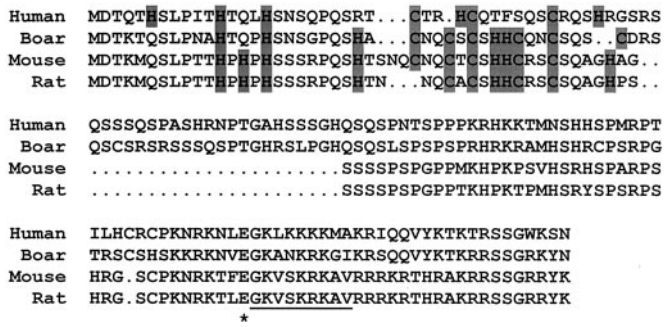


FIG. 7. **Comparison of cDNA derived amino acid sequences of human (34), boar (35), mouse (36), and rat TP2 (10) using Pileup program.** The *underlined* sequence after the glutamate (86$^{th}$ position) (*) represents nuclear localization signal. The histidine and cysteine residues that are involved in zinc coordination in rat TP2 and conserved in all the four species are highlighted.

in mammals (33). Thus the physiological significance of TP2 preferential localization to nucleolus needs to be addressed in future, particularly with respect to the events taking place in round spermatids during chromatin remodeling at the time of appearance of transition proteins.

Boar TP2 binds three atoms of zinc per molecule and hence it has been proposed to contain three fingers (23, 24). However, the finer details of the zinc finger modules in boar TP2 have not been delineated yet. A comparison of the amino acid sequences of the four mammalian TP2s (rat, mouse, boar, and human) is shown in Fig. 7. It is evident from this comparison that the N-terminal two-thirds of boar and human TP2 have diverged significantly from the rat and mouse TP2. At the same time the C-terminal one-third basic domain is fairly conserved in all the species. In fact there is a 24-amino acid stretch insertion in the boar and human TP2, which make them longer than the rat and mouse TP2. The additional histidine residues present in this insertion domain may contribute to the generation of the third finger in boar and human TP2. A more detailed analysis of the boar and human TP2 are necessary particularly with respect to the amino acid residues that are involved in the coordination of zinc. It is quite possible that the final chromatin packaging present in the sperm in rodents, boar, and humans are influenced by TP2 particularly since, both TP1 and the protamines are fairly conserved among all these species.[2] Thus, the additional zinc binding module in human and boar TP2 may therefore be relevant in contributing to chromatin packaging in these sperms. This question becomes even more important since a small percentage of mammalian sperm chromatin still retains nucleosomal structure (37, 38). At present we do not know what are the chromatin domains that retain the nucleosomal structure in human and mouse sperm. It may also be noted here that the sequence beyond the glutamate residue at position 86 (in rat) is highly conserved in all the species. Based on the results obtained here regarding the localization pattern of TP2, we would like to believe that the preferential localization of TP2 to nucleolus might be a common feature in rodents, boar, and human. Thus, in conclusion, evolutionary significance of the divergence of the zinc fingers domains of TP2 within mammals and also the physiological significance of the stage-specific appearance of these proteins during spermiogenesis only in mammals need to be understood with particular reference to mechanism of chromatin remodeling and the nature of packaged and condensed chromatin present in the sperm in these species.

[2] A. R. Meetei, K. S. Ullas, and M. R. S. Rao, unpublished observations.

REFERENCES


1. Meistrich, M. L. (1989) in *Histones and Other Basic Nuclear Proteins* (Hnilica, L. S., Stein, G. S., and Stein, J. L., eds) pp. 165–182, CRC Press, Boca Raton, FL
2. Singh, J., and Rao, M. R. S. (1987) *J. Biol. Chem.* **262,** 734–740
3. Singh, J., and Rao, M. R. S. (1988) *Biochem. Int.* **17,** 701–710
4. Yu, Y. E., Zhang, Y., Unni, E., Shirley, C. R., Deng, J. M., Russel, L. D., Weil, M. M., Behringer, R. R., and Meistrich, M. L. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 4683–4688
5. Akama, K., Kondo, M., Sato, H., and Nakano, M. (1999) *FEBS Lett.* **442,** 189–192
6. Baskaran, R., and Rao, M. R. S. (1990) *J. Biol. Chem.* **265,** 21039–21047
7. Baskaran, R., and Rao, M. R. S. (1991) *Biochem. Cell Biol.* **179,** 1491–1499
8. Kundu, T. K., and Rao, M. R. S. (1995) *Biochemistry* **34,** 5143–5150
9. Kundu, T. K., and Rao, M. R. S. (1996) *Biochemistry* **35,** 15626–15632
10. Meetei, A. R., and Rao, M. R. S. (1996) *Protein Exp. Purif.* **8,** 409–415
11. Meetei, A. R., and Rao, M. R. S. (1998) *Protein Exp. Purif.* **13,** 184–190
12. Meetei, A. R., and Rao, M. R. S. (1998) *Anal. Biochem.* **264,** 288–291
13. Laemmli, U. K. (1970) *Nature* **227,** 680–685
14. Dingwall, C., and Laskey R. A. (1991) *Trends Biochem. Sci.* **16,** 478–481
15. Kundu, T. K., and Rao, M. R. S. (1994) *FEBS Lett.* **351,** 6–10
16. Donalson, I. M., and Friesen, J. D. (2000) *J. Biol. Chem.* **275,** 13780–13788
17. Christianson, D. W. (1991) *Adv. Protein Chem.* **42,** 281–355
18. Pavletich, N. P., and Pabo, C. O. (1991) *Science* **252,** 809–817
19. Wolfe, S. A., Greisman, H. A., Ramm, E. I., and Pabo, C. O. (1999) *J. Mol. Biol.* **285,** 1917–1934
20. Klug A. (1999) *J. Mol. Biol.* **293,** 215–218
21. Pavletich, N. P., and Pabo, C. O. (1993) *Science* **261,** 1701–1707
22. Marmorstein, R., Carey, M., Ptashne, M., and Harrison, S. C. (1992) *Nature* **356,** 408–414
23. Akama, K., Sato, H., Oguma, K., and Nakano, M. (1997) *Biochem. Mol. Biol. Int.* **42,** 865–872
24. Sato, H., Akama, K., Kojima, S., Miura, K., Sekine, A., and Nakano, M. (1999) *Protein Exp. Purif.* **16,** 454–462
25. Busch, H., and Smetana, K. (1970) *The Nucleolus,* pp. 161–201, Academic Press, New York
26. Rikkonem, M., Peranen, J., and Kaariainen, L. (1992) *Virology* **189,** 462–473
27. Annilo, T., Kari, A., Hoth, S., Rikk, T., Kruppa, J., and Metspalu, A. (1998) *Biochem. Biophys. Res. Commun.* **249,** 759–766
28. Siomi, H., Shida, H., Nam, S. H., Nosaka, T., Maki, M., and Hatanaka, M. (1988) *Cell* **55,**197–209
29. Falkner, F. G., Fuerst, T. R., and Moss, B. (1988) *Virology* **164,** 450–457
30. Valdez, B. C., Peralky, S., Henning, D., Saijo, Y., Chan, P. K., and Busch, H. (1994) *J. Biol. Chem.* **269,** 23776–23783
31. Mears, W. E., Lam, V., and Rice, S. A. (1995) *J. Virol.* **69,** 935–947
32. Liu, J. L., Lee, L. F., Ye, Y., Qian, Z., and Kung, H. J. (1997) *J. Virol.* **71,** 3188–3196
33. Kundu, T. K., and Rao, M. R. S. (1999) *J. Biochem.* (*Tokyo*) **125,** 217–222
34. Nelson, J. E., and Krawetz, S. A. (1993) *J. Biol. Chem.* **268,** 2932–2936
35. Keime, S., Kumm, S., Luerssen, H., and Engel, W. (1992) *Anim. Genet.* **23,** 373–378
36. Kleene, K. C., and Flynn, J. F. (1987) *J. Biol. Chem.* **262,** 17272–17277
37. Gatewood, J. M., Cook, G. R., Bradbury, E. M., and Schmid, C. W. (1987) *Science* **236,** 962–964
38. Pittogi, C., Renzi, L., Zaccagnini, G., Cimini, D., Degrassi, F., Giordano, R., Magnano, A. R., Lorenzini, R., Lavia, P., and Spadafora, C. (1999) *J. Cell Sci.* **112,** 3537–3548