Sequence variation in transcription factor IIIA

Chris J. Gaskins and Jay S. Hanas
Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73190, USA

Received November 28, 1989; Revised and Accepted March 13, 1990

ABSTRACT

Previous studies characterized macromolecular differences between *Xenopus* and *Rana* transcription factor IIIA (TFIIIA) (Gaskins et al., 1989, Nucl. Acids Res. 17, 781–794). In the present study, cDNAs for TFIIIA from *Xenopus borealis* and *Rana catesbeiana* (American bullfrog) were cloned and sequenced in order to gain molecular insight into the structure, function, and species variation of TFIIIA and the TFIIIA-type zinc finger. *X. borealis* and *R. catesbeiana* TFIIIA have 339 and 335 amino acids respectively, 5 and 9 fewer than *X. laevis* TFIIIA. *X. borealis* TFIIIA exhibited 84% sequence homology (55 amino acid differences) with *X. laevis* TFIIIA and *R. catesbeiana* TFIIIA exhibited 63% homology (128 amino acid changes) with *X. laevis* TFIIIA. This sequence variation is not random; the C-terminal halves of these TFIIIA contain substantially more non-conservative changes than the N-terminal halves. In particular, the N-terminal region of TFIIIA (that region forming strong DNA contacts) is the most conserved and the C-terminal tail (that region involved in transcription promotion) the most divergent. Hydropathy analyses of these sequences revealed zinc finger periodicity in the N-terminal halves, extreme hydrophilicity in the C-terminal halves, and a different C-terminal tail hydropathy for *R. catesbeiana* TFIIIA. Although considerable sequence variation exists in these TFIIIA zinc fingers, the Cys/His, Tyr/Phe and Leu residues are strictly conserved between *X. laevis* and *X. borealis*. Strict conservation of only the Cys/His motif is observed between *X. laevis* and *R. catesbeiana* TFIIIA. Overall, Cys/His zinc fingers in TFIIIA are much less conserved than Cys/Cys fingers in erythroid transcription factor (Eryf 1) and also less conserved than homeo box domains in segmentation genes. The collective evidence indicates that TFIIIA evolved from a common precursor containing up to 12 finger domains which subsequently evolved at different rates.

INTRODUCTION

Transcription of 5S ribosomal RNA genes in *Xenopus* oocytes requires RNA polymerase III and at least three additional factors (1,2). One of these proteins, transcription factor IIIA (TFIIIA), has a mass of about 40,000 Daltons and promotes transcription by binding to the 50 bp internal control region (ICR) of the 5S RNA gene (1,3,4). TFIIIA contains zinc and requires the metal for function (5). The amino acid sequence of *Xenopus laevis* TFIIIA (6) contains at least nine repetitive domains (fingers), each with the potential to coordinate a zinc ion between two cysteine and two histidine residues (7,8). In vitro mutagenesis studies suggest a region comprising the N-terminal fingers has a major role in forming DNA contacts with the 3' portion of the 5S RNA gene ICR (9,10,11) whereas the C-terminal tail has a major role in promoting transcription (9). Although TFIIIA was the first DNA binding protein shown to require zinc for function and contain zinc fingers, many other gene regulatory proteins containing amino acid repeats similar to those found in TFIIIA have since been described in a wide variety of organisms (12). Types of zinc fingers found in eukaryotic gene regulatory proteins include the Cys/His motif exemplified by TfIIIA and the Cys/Cys motif exemplified by steroid hormone receptors (13) and the erythroid transcription factor (14). Zinc fingers are common evolutionary threads connecting major classes of eukaryotic DNA binding proteins. Recently, the examination of species variation in the erythroid transcription factor revealed that the Cys/Cys zinc fingers in this protein are evolving at an extremely slow rate, comparable to some histone genes (14). Earlier work documented high degrees of conservation in homeo box domains of segmentation genes (15,16). Species variation in TFIIIA was previously examined at the macromolecular level and differences were observed in the immunological, DNA binding, and transcription promotion properties of *Xenopus* and *Rana* TFIIIA (17). In the present study, the sequences of full-length, TFIIIA cDNAs from *X. borealis* and *R. catesbeiana* are reported. Similarities and differences in these amino acid sequences provide insight into TFIIIA structure and function and indicate a high rate of evolution for the TFIIIA-type zinc finger.

METHODS AND MATERIALS

RNA isolation and northern blotting

Poly(A) RNA was isolated from the ovarian tissue of immature *Xenopus borealis* (4–5 cm frogs, Nasco, WI) and *Rana catesbeiana* (6–7 cm frogs, W. Lemberger, Oshkosh, WI) by the guanidinium isothiocyanate-SDS-proteinase K method (18) utilizing the Invitrogen (La Jolla, CA) Fast Track mRNA isolation system. The tissue was homogenized in lysis buffer and the homogenate was applied directly to an oligo(dT) cellulose column for selection of poly(A) RNA. Poly(A)+ RNA was eluted off the column in low ionic strength buffer, ethanol precipitated, and quantitated by absorbance at 260 nm. Approximately 100 µg of
Isolation of cDNA inserts and DNA sequence analysis

Lambda gt10 bacteriophage clones containing cDNA inserts that cross-hybridized with the X. laevis TFIIIA cDNA were purified as described previously (23). Phage DNAs were digested with EcoRI or NotI and electrophoresed through a 1% agarose gel. Insert cDNAs were electroeluted and ligated into Phagemid M13 bacteriophage RF (Stratagene, La Jolla, CA) that had been digested with EcoRI or NotI. Single-stranded phage DNA containing inserts were purified and partially sequenced by the dideoxy chain termination method (24) using a modified form of T7 DNA polymerase (Sequenase, United States Biochemical, Cleveland, OH). One cDNA insert was sequenced completely on both strands utilizing a series of complementary oligonucleotides as primers (obtained from K. Jackson, St. Francis Research Facility of Tulsa, OK). Nucleotide and amino acid sequences were analyzed with software packages from Genetics Computer Group (Madison, WI, ref. 25) and Beckman Instruments (Palo Alto, CA, ref. 26).

RESULTS

Identification of TFIIIA mRNA from X. borealis and R. catesbeiana

Ovarian tissue RNA from immature X. borealis and R. catesbeiana was examined for the presence of transcripts which cross-hybridize with the 32P-labeled Xenopus laevis TFIIIA cDNA (6). Fig. 1 is a northern blot of X. laevis liver RNA (lane 1) or ovarian tissue RNA (lane 2), X. borealis ovarian tissue RNA that had been poly(A)+ selected (lane 3), and R. catesbeiana poly(A)+ RNA (lane 4) all probed with X. laevis TFIIIA cDNA. X. borealis ovarian tissue contains a major poly(A)+ RNA species (lane 3) migrating in similar fashion to the 1.5 kb X. laevis TFIIIA mRNA present in total ovarian RNA (lane 2). R. catesbeiana ovarian tissue also contains a similarly migrating mRNA that cross-hybridizes with this TFIIIA probe (lane 4). The X. laevis TFIIIA probe did not detect under these conditions any cross-hybridizing RNA species in X. laevis liver (lane 1) because of the high selectivity of this probe and the extremely low abundance of TFIIIA mRNA in this organ compared to ovarian tissue. The greater abundance of the TFIIIA mRNA in the X. borealis ovarian RNA compared with X. laevis ovarian RNA is due to poly(A)+ enrichment. The weakly hybridizing R. catesbeiana mRNA is most likely due to sequence diversity.

Sequence of X. borealis and R. catesbeiana TFIIIA cDNA

cDNA libraries were constructed from the same preparations of X. borealis or R. catesbeiana poly(A)+ RNA electrophoresed in Fig. 1. After screening about 1 x 107 lambda gt10 recombinants from each library with the X. laevis TFIIIA cDNA probe, seven positive X. borealis clones and two positive R. catesbeiana clones were isolated. Four of the X. borealis clones and two of the R. catesbeiana clones contained inserts of about 1.4 kb, the approximate size of the respective TFIIIA mRNAs (Fig. 1). One clone from each species was chosen for complete sequencing on both strands using a series of complementary oligonucleotides as sequencing primers. The sequences of the non-coding strands (minus the majority of the poly(A) tail) and the predicted amino acid sequence (numbered on left margin) are given in Figs. 2 and 3. The X. borealis cDNA (Fig. 2) consists of 1363 bp plus a EcoRI-NotI linker at the 5' end and a poly(A) tail. An open reading frame, coding for a protein of 339 amino acids, begins at position 67 and extends to a TAA termination codon at position...
The nucleotide and deduced amino acid sequences of *Xenopus borealis* TFIIIA cDNA. cDNA synthesis, cloning, and sequencing were performed as described in METHODS AND MATERIALS. Partial sequences of two other independent clones were in complete agreement with this sequence. A putative polyadenylation signal is located at nucleotides 1359-1364.
Fig. 3. The nucleotide and deduced amino acid sequence of *Rana catesbeiana* TFIIIA cDNA. cDNA synthesis, cloning, and sequencing were performed as described in the METHODS. A partial sequence from one other independent clone was in complete agreement with this sequence. A putative polyadenylation signal is located at nucleotides 1242-1247.
1084. The nucleotide sequence around the ATG start codon at nucleotide position 67 shares a low level of homology with a eukaryotic translation initiation consensus sequence (GCCGCCRCCAUGG) described by Kozak (27). A somewhat similar sequence is found in the X. laevis TFIIA mRNA (6).

The 3' untranslated sequence contains a near canonical polyadenylation signal (ATTAAA) at position 1359. The R. catesbeiana TFIIIA cDNA (Fig. 3) contains 1249 bp plus a NotI linker at the 5' end and a poly(A) tail. An open reading frame coding for a protein of 335 amino acids commences at position 20 and terminates at position 1025. The ATG region of the R. catesbeiana TFIIIA cDNA also shares low homology with the translation initiation consensus sequence. A potential polyadenylation signal is found at nucleotide 1242 (underlined). The codon usages of X. laevis, X. borealis, and R. catesbeiana TFIIIA are similar (data not shown).

Fig. 4 depicts the amino acid sequences of the X. borealis TFIIIA (A) and R. catesbeiana TFIIIA (B) folded into the zinc finger pattern proposed for X. laevis TFIIIA by Tso et al. (28). Overall, the X. borealis protein (A) has 84% amino acid sequence homology with X. laevis TFIIIA, differing at 55 residues and R. catesbeiana TFIIIA (B) shares 63% amino acid sequence homology with X. laevis TFIIIA, with 128 amino acid differences. The most conserved and diverged regions are the N-terminal two fingers and the C-terminal tail respectively. The X. laevis and X. borealis zinc fingers conserve the zinc-coordinating cysteines and histidines and aromatic/hydrophobic amino acids in the proper position consistent with the zinc finger consensus sequence (Phe/Tyr-X-Cys-X2 or 4-Cys-X3-Phe-X5-Leu-X2-His-X3-His-X5) as described by Berg (29). Although the Cys/His amino acids in this consensus zinc finger motif are conserved between X. laevis and R. catesbeiana, the first Phe in finger 3 (now a Cys) and the Leu in finger 6 (now a Cys) of R. catesbeiana TFIIIA are not conserved. Relative to X. laevis TFIIIA, both the X. borealis and R. catesbeiana TFIIIAs have fewer amino acids in their C-terminal tails (dark circles, Fig. 4). A valine residue at position 313 of X. laevis TFIIIA is deleted in X. borealis TFIIIA. In addition, the four C-terminal residues of X. laevis TFIIIA are deleted in the X. borealis TFIIIA. Amino acid deletions are much more extensive in the C-terminal tail of R. catesbeiana TFIIIA (Fig. 4B). Although 13 amino acids are missing in the tail region, overall the R. catesbeiana protein contains 9 fewer amino acids because of an insertion of 4 amino acids between the 6th and 7th zinc fingers. Table I lists the numbers of conservative and non-conservative amino acid changes (relative to X. laevis TFIIIA) in the N-terminal and C-terminal halves of X. borealis and R. catesbeiana TFIIIA.

Moreover, the distribution of amino acid changes is not random. Although the C-terminal halves of the proteins do have increased numbers of conservative changes compared to the N-terminal halves, a much larger percent increase in non-conservative changes is observed in the C-terminal halves. These data support a previous prediction that the C-terminal half of Rana TFIIIA has considerably more sequence variation than the same region of Xenopus TFIIIA (17).

Hydropathy analysis is a direct way of comparing primary structures of related proteins and an indirect predictor of tertiary
Table 1. Distribution of amino acid changes in TFIIIA*

<table>
<thead>
<tr>
<th></th>
<th>X. borealis</th>
<th>R. catesbeiana</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>conservative</td>
<td>non-conservative</td>
</tr>
<tr>
<td>N-terminal half</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>C-terminal half</td>
<td>11</td>
<td>26</td>
</tr>
<tr>
<td>% increase in C-terminal half</td>
<td>+136</td>
<td>+31</td>
</tr>
</tbody>
</table>

*relative to X. laevis TFIIIA

structure. Fig. 5 exhibits the hydropathy plots of X. laevis (A), X. borealis (B), and R. catesbeiana (C) TFIIIA. Overall, the TFIIIA primary structures are very hydrophilic (− hydropathy values) which is compatible with an extended tertiary structure in an aqueous environment. The C-terminal halves, possessing large hydrophilic valleys centered at about amino acid 215 and 285, are more hydrophilic than the N-terminal halves. This observation is consistent with other data indicating structural differences between the N-terminal and C-terminal halves of TFIIIA (30). A distinct hydropathy periodicity corresponding to zinc fingers 1—6 is observed in the first 190 amino acids. Exon periodicity corresponding to zinc fingers 1—6 exists in the TFIIIA gene as well (28). Hydropathy differences do exist between the primary structures of these TFIIIA species. Zinc fingers 2 and 3 (peaks centered at about amino acids 55 and 82 in Fig. 5) are most hydrophobic in R. catesbeiana TFIIIA whereas zinc finger 5 (peak centered at about amino acid 148) is most hydrophobic in X. borealis TFIIIA. The linker region between zinc fingers 6 and 7 (centered at about amino acid 196) is most hydrophobic in R. catesbeiana TFIIIA. The hydropathy profiles of the C-terminal tail regions between amino acids 310 and 330 of X. borealis and X. laevis TFIIIA are very similar whereas this profile is different in R. catesbeiana TFIIIA.

DISCUSSION

Species variation between Xenopus and Rana TFIIIA was previously studied at the macromolecular level (17). Results from this earlier study indicated 1) R. catesbeiana TFIIIA had a slightly greater electrophoretic mobility than X. borealis TFIIIA during SDS PAGE (mobility of X. borealis TFIIIA was slightly greater than X. laevis TFIIIA), 2) a rabbit polyclonal antiserum principally against the C-terminal half of X. laevis TFIIIA reacted strongly with X. laevis and X. borealis TFIIIA in an immunoblot but only marginally with R. catesbeiana TFIIIA, 3) whereas X. laevis or X. borealis TFIIIA protected from DNase I digestion the entire Xenopus 5S gene ICR (+96 to +43), R. catesbeiana TFIIIA only afforded complete protection from +96 to about +78 on the Xenopus 5S gene ICR, 4) R. catesbeiana TFIIIA was less efficient than X. laevis or X. borealis TFIIIA at promoting transcription of the Xenopus 5S gene in vitro. In the present study, cDNA clones for TFIIIA from X. borealis and R. catesbeiana were isolated and sequenced to elucidate structural differences at the molecular level between Xenopus and Rana TFIIIA and to better understand evolution of TFIIIA and the Cys/His type zinc finger. The slight differences in molecular weight observed between these TFIIIA as SDS PAGE most likely reflect the differing numbers of amino acids encoded by their cDNAs. The marginal binding to R. catesbeiana TFIIIA by a polyclonal antiserum directed principally against the C-terminal half of Xenopus TFIIIA could be caused by the large number of non-conservative amino acid changes in the C-terminal half of the Rana protein (Table I). R. catesbeiana TFIIIA or X. laevis TFIIIA with the fourth zinc finger disrupted both footprint the Xenopus 5S RNA gene in a very similar fashion i.e., strong DNase I protection by that region of the protein comprising the N-terminal fingers (11,17). Compared to X. laevis and X. borealis TFIIIA, R. catesbeiana TFIIIA contains a large number of amino acid substitutions in the finger 3—4 region which may account for this DNase I protection pattern on the Xenopus 5S RNA gene. A comparison between DNase I protection by X. laevis or X. borealis TFIIIA on the coding and non-coding strands of the X. borealis somatic 5S RNA gene revealed near identical patterns except in the +63 region (new hypersensitive site on coding strand and shift in hypersensitive site on non-coding strand, ref. 17). According to the model of Tso et al.(28), finger 7 is most

![Fig. 5. Hydropathy plots of Xenopus and Rana TFIIIA. The hydropathy analysis was performed with a VAX computer using the algorithm of Kyte and Doolittle (33) via the GCG software package (25). The hydrophobic index (ranging from 2 to −4, most hydrophobic to most hydrophilic) is plotted versus TFIIIA amino acid sequence. A, X. laevis TFIIIA; B, X. borealis TFIIIA; C, R. catesbeiana TFIIIA.](image)
likely interacting with this region of the 5S gene ICR. It is noted that *X. borealis* finger 7 has twice as many amino acid changes (compared with *X. laevis* TFIIIA) than other fingers. The reduced transcription efficiency of *R. catesbeiana* TFIIIA on the *Xenopus* 5S RNA gene compared to *X. laevis* or *X. borealis* TFIIIA could be due to decreased binding affinity and/or structural alterations in the C-terminal tail. In that region of the C-terminal tail of *X. laevis* TFIIIA previously shown to be absolutely required for transcription promotion (amino acids 295–313, ref. 9), *R. catesbeiana* TFIIIA has 1 conservative and 7 non-conservative changes plus 2 deletions whereas *X. borealis* TFIIIA has only 2 conservative and 4 non-conservative changes and no deletions.

The overall amino acid sequence of TFIIIA is not highly conserved. There exists 84% amino acid sequence homology (89% DNA sequence homology) between *X. laevis* and *X. borealis* TFIIIA and 63% homology (56% DNA sequence homology) between *X. laevis* and *R. catesbeiana* TFIIIA. These are the first reported TFIIIA sequences from other species since the original report on *X. laevis* TFIIIA (6); the amino acid and DNA sequence homology in TFIIIA within the *X. laevis* species was found to be 98% and 99% (28). The C-terminal half of *R. catesbeiana* TFIIIA has many more non-conservative changes relative to *X. laevis* TFIIIA (Table I). This observation is consistent with other data indicating structural differences between the N-terminal and C-terminal halves of *Xenopus* and *Rana* TFIIIA (17). Although individual TFIIIA zinc fingers have widely varying sequences, the Cys/His motif is strictly conserved between *Xenopus* and *Rana*. Less than strict conservation is observed in the aromatic/hydrophobic residues of the TFIIIA finger motif. The N-terminal region (first two zinc fingers) has less sequence variation than other regions of TFIIIA possibly due to the importance of this region in forming strong contacts with the 3′ portion of the 5S gene ICR (9,10,11). The His-Thr-Gly-Glu-Lys pentapeptide found exclusively in fingers 1 and 2 is conserved (Arg is substituted for Lys in *Rana* finger 1). This pentamer is the largest conserved amino acid string in TFIIIA. It is also found in other Cys/His finger proteins, e.g. Sp1 (31).

The sequence variation of TFIIIA contrasts in interesting and informative ways with a similar study recently reported on the erythroid transcription factor, Eryf 1, (14). Eryf 1, which binds promoters and enhancers in the globin gene family, contains related N-terminal and C-terminal domains plus two internal zinc fingers of the Cys/Cys motif. The two N-terminal repeats (199 a.a.) of human and mouse Eryf 1 have 13 conservative and 25 non-conservative amino acid differences and the C-terminal repeat (103 a.a.) has 3 conservative and 11 non-conservative changes. In contrast, the zinc finger domains (108 a.a) have just 2 conservative amino acid differences; finger domains of chicken Eryf 1 have just 3 conservative and 5 non-conservative changes (no Cys/Cys changes) relative to the human protein (14). These observations led to the conclusion that Eryf 1 evolved from a precursor containing two distinct classes of domains (DNA or protein binding) which subsequently evolved at greatly different rates. The high conservation (similar to histones) of the DNA binding, Cys/Cys finger domains suggested greater constraints existed on protein-DNA recognition than protein-protein interactions. Evolution of the DNA and protein binding domains of TFIIIA may have progressed from a common precursor containing as many as 12 finger domains. This number of repeats is inferred from a computer study of TFIIIA in which 12 repeats were observed, with considerably more variation in the C-terminal 3 repeats (8).

**ACKNOWLEDGEMENTS**

This study was supported in part by grants from the National Institute of General Medical Sciences and the National Science Foundation. The authors thank R. Denell for helpful discussions and B. Roe for assistance with the VAX computer.

**REFERENCES**