# Contribution of Individual Base Pairs to the Interaction of TFIIIA with the *Xenopus* 5S RNA Gene†

Nik Veldhoen,‡ Qimin You,‡ David R. Setzer,§ and Paul J. Romaniuk*,‡

*Department of Biochemistry and Microbiology, University of Victoria, P.O. Box 3055, Victoria, British Columbia, Canada V8W 3P6, and Department of Molecular Biology and Microbiology, School of Medicine, Case Western Reserve University, Cleveland, Ohio 44106*

*Received January 19, 1994; Revised Manuscript Received April 8, 1994®*

ABSTRACT: The effects of a series of point mutations within the *Xenopus borealis* somatic-type 5S RNA gene on transcription factor IIIA (TFIIIA) binding affinity were quantified. These data define a critical sequence-dependent contact region within the classical box C promoter element from base pair 80 to 91. Substitution of GC base pairs at positions 81, 85, 86, 89, and 91 significantly reduce TFIIIA binding affinity. Base pairs located at other positions within the box C contact region provide a moderate contribution to TFIIIA–5S gene interaction. In contrast to the extensive set of sequence contacts within the box C element, TFIIIA interaction is localized primarily to two GC base pairs at positions 70 and 71 within the intermediate promoter element. A selected amplification and binding assay (SAAB) was performed with a synthetic internal control region (ICR) randomized from base pair 78 to 95 to identify box C promoter sequences bound with high affinity by TFIIIA. The wild-type 5S RNA gene sequence from 79 to 92 is strongly selected. These results are consistent with the critical role of the box C element in sequence-dependent promoter recognition by TFIIIA.

Transcription of the *Xenopus* 5S rRNA genes by RNA polymerase III requires the formation of a stable preinitiation complex (Bogenhagen et al., 1980). An early event in the assembly of this ternary complex is the binding of transcription factor IIIA (TFIIIA)[1] to the internal control region of the 5S RNA gene (Engelke et al., 1980; Sakonju & Brown, 1981). This metastable complex is subsequently bound by TFIIIC and TFIIIB (Bogenhagen et al., 1982; Lasser et al., 1983). RNA polymerase III recognizes this preinitiation complex and is aligned on the gene in the correct orientation to initiate transcription from a start site approximately 50 base pairs upstream. This transcription complex remains bound to the DNA through multiple passages of the RNA polymerase (Bogenhagen et al., 1982).

TFIIIA is a unique member of the "zinc finger" family of transcription factors, as it exhibits specific binding to both DNA (the 5S rRNA gene) and 5S RNA, forming a 7S RNP storage particle in the cytoplasm of the immature *Xenopus* oocyte (Pelham & Brown, 1980; Picard et al., 1980). This 40-kilodalton (kDa) protein contains a 30-kDa nucleic acid binding region of nine zinc finger motifs and a 7–8-kDa activation domain (Miller et al., 1985; Brown et al., 1985; Smith et al., 1984). Each zinc finger domain consists of about thirty amino acids, with highly conserved pairs of cysteine and histidine residues binding a single zinc ion (Diakun et al., 1986; Berg, 1988). On the basis of analyses of model peptides containing one to three zinc fingers, the tertiary structure of each finger is likely to be stabilized by hydrophobic interactions between conserved nonpolar residues. Zinc fingers in the *zif*268 protein, which may be structurally similar to those in TFIIIA, interact with the DNA sequence in the major groove using a short amphipathic $\alpha$-helix located at the carboxyl terminus of the zinc finger (Pavletich & Pabo, 1991).

Studies using deletion mutants initially defined the minimal sequence of the 5S RNA gene required for transcription initiation to be base pairs 50–83 (where 1 denotes the site of transcription initiation) (Sakonju et al., 1980; Bogenhagen et al., 1980). Similar studies using a series of point mutants extended this internal control region (ICR) from base pair 45 to 97 (Pieler et al., 1985a). The ICR is tripartite in structure, consisting of box A (50–64) and box C (80–97) elements flanking a small intermediate element (67–72) (Pieler et al., 1985b, 1987; You et al., 1991). Spacer regions between these promoter elements are quite tolerant of changes in sequence (Pieler et al., 1985b). The box A sequence is homologous to the box A or D control region of tRNA genes (Ciliberto et al., 1983), while the intermediate and box C elements are specific to the 5S RNA gene. Similarity between promoter elements of 5S RNA and tRNA genes is not surprising, as they share common transcription factors (Lasser et al., 1983; Shastry et al., 1982).

Point mutations within the box C region of the *Xenopus* 5S RNA gene severely affect the assembly of a functional transcription complex, while changes in the sequence of box A or the intermediate element result in moderate effects on transcription (Pieler et al., 1985a,b). These latter two elements may function in correctly aligning the bound transcription factors on 5S DNA. Four point mutations were tested for direct effects on TFIIIA binding; the results demonstrated a critical role for the box C element in promoter recognition (Pieler et al., 1987).

Experiments using a bacterially expressed polypeptide containing the first three fingers of TFIIIA showed that the three-finger peptide could bind to base pairs 80–92 of the 5S

† This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (OGP0037673) to P.J.R. and the National Science Foundation (DMB 88-19286) and National Institutes of Health (GM-48035) to D.R.S.

* Corresponding author.
‡ Univeristy of Victoria.
§ Case Western Reserve University.

® Abstract published in *Advance ACS Abstracts,* May 15, 1994.

[1] Abbreviations: TFIIIA, transcription factor IIIA; ICR, internal control region; SAAB, selected amplification and binding; PCR, polymerase chain reaction.

0006-2960/94/0433-7568$04.50/0 

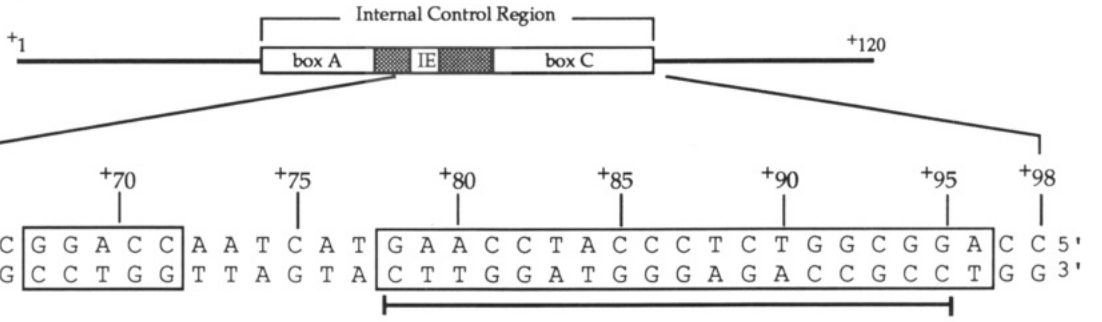



FIGURE 1: Location of point mutations and randomized base pairs within the 5S RNA gene internal control region. A series of point mutations within the intermediate and box C promoter elements analyzed for TFIIIA binding affinity are shown as boxed sequence. The sequence randomized within a synthetic ICR for SAAB analysis is underlined. Position 1 denotes the start site of transcription.

RNA gene with an affinity similar to that observed with intact TFIIIA (Christensen et al., 1991; Liao et al., 1992). These results supported earlier studies using truncated TFIIIA mutants produced in a cell-free translation system (Vrana et al., 1988). Recently, "broken finger" mutants of TFIIIA, in which a coordinating histidine is replaced with asparagine, exhibit a more complex pattern of zinc finger contribution to the DNA binding energy (Del Rio et al., 1993). Finger 3 contributes a large part toward complex formation, while disruption of fingers 1 or 2 had moderate effects on the free energy of binding. Thus, current experimental evidence is consistent with the existence of multiple energetically favorable contacts between TFIIIA and the 5S gene within the box C promoter element.

The present study investigates the contribution of individual base pairs within box C to the overall interaction of *Xenopus* TFIIIA with the *Xenopus borealis* somatic-type 5S RNA gene. The effects of point mutations within the intermediate and box C elements (Figure 1) on the equilibrium association constant for TFIIIA were quantified using a nitrocellulose filter binding assay. In addition, a selected amplification and binding (SAAB) assay (Blackwell & Weintraub, 1990) was performed to identify box C sequences, from positions 78 to 95, that bind with high affinity to TFIIIA. Sequence-specific interactions between the transcription factor and the box C promoter element extend from base pairs 80 to 91. Every base pair within this region contributes to promoter recognition and binding by TFIIIA. Base pairs flanking the box C promoter element that do not contribute to the DNA binding energy may also influence complex formation. Thus, a combination of local DNA conformation and specific base sequence interactions contributes to box C element–TFIIIA association.

## MATERIALS AND METHODS

*Bacterial Strains and Vectors.* Plasmid pT7-TF, which contains the TFIIIA cDNA from pUC 3a1.b (Ginsberg et al., 1984) cloned into the *Nde*I/*Bam*HI sites of pT7-7, was provided by J. Tso. The *Nde*I/*Bam*HI fragment of pT7-TF was cloned into the same site of pET-11b (Studier et al., 1990) to yield pTF4.

The phagemid pST5RD contains the *Xenopus borealis* somatic-type 5S RNA gene from pXbs201 cloned into the *Hin*cII site of pGP76. Construction of this clone and oligonucleotide-directed mutagenesis used to generate the series of point mutants used in this study have been described previously (Setzer et al., 1990).

DNA vectors were maintained in *Escherichia coli* K-12 strain NM522. TFIIIA expression from pTF4 was carried out using the *E. coli* B strain BL21(DE3) (F′ompTr$^{-}_{B}$m$^{-}_{B}$).

*Purification of Recombinant TFIIIA.* Expression and purification of recombinant *Xenopus* TFIIIA were performed as described by Del Rio and Setzer (1991) with the following modifications: TFIIIA was eluted from the Bio-Rex 70 column (Bio-Rad) by step elution (400 mM NaCl followed by 600 mM NaCl in buffer A/5 M urea) and used directly for DNA binding assays. This TFIIIA preparation is approximately 85% pure as estimated by SDS–PAGE. Protein concentration was determined as described previously (Romaniuk, 1985).

*PCR-Based Labeling of the 5S RNA Gene.* The 5S rRNA gene was internally labeled in a 10-$\mu$L PCR reaction mixture containing 1× PCR buffer (10 mM Tris-HCl, pH 8.3, at 20 °C, 1.5 mM $MgCl_2$, and 25 mM KCl), 50 $\mu$g/mL gelatin, 1 ng of template DNA, 150 pmol each of dTTP, dGTP, and dCTP, 5 pmol of dATP, 6.6 pmol (20 $\mu$Ci) of [$\alpha$-$^{32}$P]dATP, 0.5 pmol of primers (550 and 632), and 0.25 unit of Taq DNA polymerase. Twenty-five rounds of thermal cycling were carried out using a denaturation temperature of 94 °C, an annealing temperature of 45 °C, and an extension temperature of 72 °C. The coding template strand of Xbs 5S DNA was primed by oligomer 550 (5′CCCCCAGAAGGCAGCA-CAAG3′); its 5′ end corresponds to position –45 upstream of the transcription start site. Oligomer 632 (5′AAGCCTAC-GACACCTGGT3′) anneals to the noncoding template strand of 5S DNA; its 5′ end corresponds to position 120. Labeled DNA products were purified as described in Romaniuk (1990).

*Equilibrium Binding of TFIIIA to Mutant 5S RNA Genes.* The equilibrium constants for the binding of radioactively labeled mutant 5S RNA genes to TFIIIA were determined by using a nitrocellulose filter binding assay (Romaniuk, 1990). A minimum of three independent determinations were carried out, using the wild-type *Xenopus borealis* somatic gene as a control for normalization in each experiment. The free energy of TFIIIA–5S RNA gene complex formation was determined from the apparent association constant using the standard Gibbs free energy equation: $\Delta G° = -RT \ln(K_a)$.

*Selected Amplification and Binding (SAAB) Assay.* The SAAB assay was carried out as described by Blackwell and Weintraub (1990) with several minor modifications. The 91 nucleotide long template consisted of the noncoding strand of the 5S RNA gene from base pair 55 to 99, with the sequence of base pairs 78–95 randomized. Unique restriction sites (underlined below) were incorporated immediately upstream and downstream of the *Xenopus* sequence. The 5′ end of the template incorporated the sequence of the M13 universal forward sequencing primer, while the 3′ end of the template consisted of the sequence complementary to the M13 universal reverse sequencing primer.

+55

5'GTTTTCCCAGTCACGACGAATTCTACAGGGTCGGGCCTGGTTAGTANNNN

F.U.P. EcoRI

+99

NNNNNNNNNNNNNNTGGGCTGCAGGTCATAGCTGTTTCCTG3'

PstI R.U.P. (comp.)

This template oligonucleotide was converted into a radioactively labeled double-stranded DNA as described (Blackwell & Weintraub, 1990) using the M13 universal primers. In the first five selection rounds, TFIIIA concentration was 100 nM and DNA bound to TFIIIA was selected by nitrocellulose filter binding. Selected DNA was recovered by phenol–chloroform extraction of the filter. In the next three rounds of selection, the TFIIIA concentration was reduced to 25, 3, and 1.5 nM, respectively. DNA bound to TFIIIA was isolated using a gel retention assay as described (Blackwell & Weintraub, 1990). Finally, DNA was isolated from the selected pool using 5 nM TFIIIA and cloned into pUC19 for sequencing.

## RESULTS

The nitrocellulose filter binding assay was used to determine the TFIIIA binding affinities for mutant 5S RNA genes. This method identifies base pair positions that contribute to the energy of the 5S RNA gene–TFIIIA association. Each position on the DNA may influence TFIIIA binding by providing an optimal stereospecific conformation or direct base contact. The severity on TFIIIA binding affinity of all substitutions at some base pair positions is consistent with the loss of direct base contacts necessary for protein binding. However, mutations at other base pair positions lead to more moderate reductions in TFIIIA association, suggesting either weaker forms of direct contact or subtle DNA conformational effects on binary complex formation.

Base pair substitutions from position 67 to 71 within the intermediate element were assayed for TFIIIA binding affinity (Table 1). All substitutions at GC70 significantly reduced TFIIIA binding affinity, suggesting that this position might provide a direct contact site for the protein within the central region of the promoter. Changing the base pair identity at GC71 had a modest influence on the TFIIIA–5S RNA gene interaction. Substitution mutations at positions CG67 to TA69 did not affect TFIIIA binding affinity.

An extensive region within box C is important for high-affinity binding of TFIIIA to the 5S RNA gene (Table 2). Substitution mutations from position 80 to 91 reduced TFIIIA binding affinity to varying degrees. Specific sequences within box C that flank this contact region (78–79 and 92–96) are not required for high-affinity interactions with TFIIIA. Different base pair substitutions at each position within the central region of box C had different effects on TFIIIA binding affinity, which may result from sequence-dependent changes in a DNA conformation optimal for binding or the loss of contact sites critical for complex formation. The diversity of responses to base pair substitution is perhaps best illustrated by comparing effects at positions with the same wild-type base pair. For example, mutation of TA80 or TA84 to a GC base pair results in similar effects on TFIIIA interaction; substitution of a CG base pair at position 84 leads to a 70% reduction in binding affinity, whereas the same change at position 80 results in no significant alteration in the binding constant (Table 2). The differential effects of substitution mutations can also be seen for AT base pairs at positions 83, 88, and 90. Substitution to GC or TA base pairs at these three sites affects TFIIIA binding to varying degrees. The clearest example of nonequivalent contribution to TFIIIA association can be seen for GC base pairs within the box C contact region. The severity of the effects of all substitution mutations at positions GC81, GC85, GC89, and CG91 strongly supports the existence of a role for these base pairs in providing direct base contacts to TFIIIA (Table 2). In contrast, the effects on TFIIIA binding affinity of different substitutions at GC82, GC86, and GC87 vary up to 3-fold and are not the same for each base pair position. Perhaps these three GC base pairs influence TFIIIA binding affinity by contributing to an optimal DNA conformation within box C.

Sequence specificity of TFIIIA binding was also analyzed using a SAAB assay. An ICR template pool, containing randomized sequence within the box C element from position 78 to 95, was subjected to repeated rounds of TFIIIA selection and amplification. As a group, the DNAs selected for high-affinity binding to TFIIIA have a consensus sequence from 80 to 92 identical to that of the *Xenopus* wild-type 5S gene sequence (Figure 2). The selected box C consensus sequence also shows strong similarity to the eukaryotic 5S DNA consensus sequence identified by Erdmann et al. (1985). The

Table 1: Effects on TFIIIA Binding Affinity[a] of 5S RNA Gene Point Mutations within the Intermediate Promoter Element

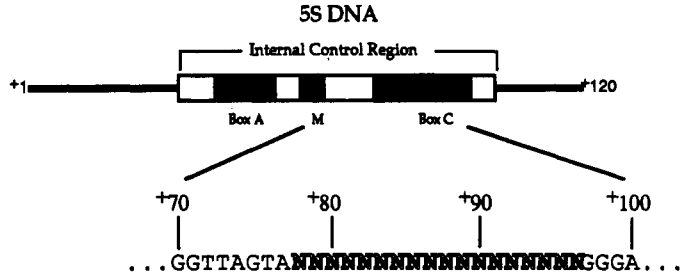
| base position | base pair substitution | | | |
|---|---|---|---|---|
| | AT | CG | GC | TA |
| CG67 | 0.89 ± 0.07 | | 1.08 ± 0.20 | 1.09 ± 0.18 |
| CG68 | 1.04 ± 0.06 | | 1.05 ± 0.14 | 1.07 ± 0.03 |
| TA69 | 0.92 ± 0.03 | 0.95 ± 0.15 | 0.98 ± 0.04 | |
| GC70 | 0.31 ± 0.09 | 0.25 ± 0.04 | | 0.67 ± 0.04 |
| GC71 | 0.64 ± 0.11 | 0.74 ± 0.09 | | 0.82 ± 0.11 |

[a] Determined as the ratio of the apparent association constant for the mutant nucleic acid to the apparent association constant for the wild-type nucleic acid. Values are averages of three or more independent determinations.

Table 2: Effects on TFIIIA Binding Affinity[a] of 5S RNA Gene Point Mutations within the Box C Promoter Element

| base position | base pair substitution | | | |
|---|---|---|---|---|
| | AT | CG | GC | TA |
| CG78 | 1.11 ± 0.25 | | 1.22 ± 0.10 | 1.00 ± 0.20 |
| TA79 | 1.12 ± 0.13 | 1.12 ± 0.13 | 1.09 ± 0.21 | |
| TA80 | 0.44 ± 0.17 | 0.98 ± 0.02 | 0.84 ± 0.04 | |
| GC81 | 0.30 ± 0.06 | 0.13 ± 0.03 | | 0.22 ± 0.05 |
| GC82 | 0.40 ± 0.01 | 0.54 ± 0.08 | | 0.90 ± 0.10 |
| AT83 | | 0.70 ± 0.07 | 0.58 ± 0.04 | 0.20 ± 0.09 |
| TA84 | 0.57 ± 0.13 | 0.39 ± 0.10 | 0.82 ± 0.14 | |
| GC85 | 0.23 ± 0.07 | 0.23 ± 0.15 | | 0.19 ± 0.07 |
| GC86 | 0.75 ± 0.18 | 0.35 ± 0.14 | | 0.43 ± 0.22 |
| GC87 | 0.49 ± 0.08 | 0.61 ± 0.19 | | 0.56 ± 0.07 |
| AT88 | | ND[b] | 0.70 ± 0.06 | 0.39 ± 0.07 |
| GC89 | ND | 0.19 ± 0.06 | | 0.25 ± 0.08 |
| AT90 | | 0.90 ± 0.10 | 0.40 ± 0.02 | 0.50 ± 0.06 |
| CG91 | 0.48 ± 0.06 | | 0.24 ± 0.05 | 0.39 ± 0.13 |
| CG92 | 0.99 ± 0.22 | | 1.27 ± 0.19 | 1.07 ± 0.04 |
| GC93 | 1.34 ± 0.14 | 1.14 ± 0.16 | | 1.06 ± 0.10 |
| CG94 | 1.02 ± 0.19 | | 1.13 ± 0.09 | 1.20 ± 0.25 |
| CG95 | 1.14 ± 0.04 | | 1.06 ± 0.05 | 1.07 ± 0.05 |
| TA96 | 0.90 ± 0.14 | 0.92 ± 0.08 | 0.94 ± 0.10 | |

[a] Determined as the ratio of the apparent association constant for the mutant nucleic acid to the apparent association constant for the wild-type nucleic acid. Values are averages of three or more independent determinations. [b] ND, not determined.




| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AGTGGATGGCTGACAACT | 1 | GTTGGAGGGGAGACCGTC | 26 |
| GATGGATGGGAGACTACA | 2 | GCAGGATGGGAGACTAGT | 27 |
| TCCGGATGGGTATCCCAA | 3 | ACTGGATGGATCCCTTAT | 28 |
| ACGGGATGGGAATGCATT | 4 | CATGGATGGGATACCGTG | 29 |
| CCTGGATGG-AGACCGCCC | 5 | TTTGGATGGGGTCCACGC | 30 |
| ATTGGATGAGGAGTATAA | 6 | GCTGGATGGGAGACTGGA | 31 |
| TTCGGATGGACCTCGTCT | 7 | TTCGGATGGGTATCCAAT | 32 |
| GTTGGAGGGGAGACCGTC | 8 | CGTGGATGACAGATGTCG | 33 |
| TACGGATGGGAGACGCCC | 9 | ACTGGATGGGAGACCTAA | 34 |
| ATTGGATGG-AGTCTTCCT | 10 | GATGGATGGGTACTGCCT | 35 |
| TCTGGATGGTGTACTGCC | 11 | CTGGGATGGATGACCAGT | 36 |
| GCCGGATGGGAGACCGAT | 12 | TTGGGAGGTTGCACTCCC | 37 |
| TCTCGGATGAGGGTACGG | 13 | GCCGTATGGGAGACCGCC | 38 |
| AACGAATGGTGAACCTCA | 14 | ATGCGGATGCTGGGTCCA | 39 |
| AATGGATGGGAGACGCCT | 15 | TTGGATGAGAGACGGGTA | 40 |
| TTCGGATGGACCTCGTCT | 16 | AATTGGAGGGGAAACTAA | 41 |
| GCCGGATGGAGGGCAAGT | 17 | AGTGGATGGTGATTACTG | 42 |
| TCTGGATGGTGTACTGCC | 18 | AATCGGATGGGAGACCTT | 43 |
| AATTGGAGGGGAAACTAA | 19 | GTTGGAGGGGAGACCGTC | 44 |
| ACTGGA-GGGTGGACGGTT | 20 | GGTGGATGGGAGACTCCC | 45 |
| TTTGGACTGATGGACGCC | 21 | ACTGGATGGGAGACCTAA | 46 |
| ACCTGGATGATGTACATA | 22 | AATGGATGGGAGACGCCT | 47 |
| GATGGATGGGTACTGCCT | 23 | TCCGGATGGGTATCCCAA | 48 |
| TATGGATGGGATGCGACA | 24 | ACGGGATGGGAATGCATT | 49 |
| TTTGGA-GGGAGTCCAGGG | 25 | ATTGGATGGGAGACCGCC | 50 |

| | Sequence Position | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 |
| Base Occurrence A | 17 | 9 | 1 | 0 | 1 | 50 | 0 | 0 | 5 | 6 | 27 | 12 | 29 | 3 | 6 | 12 | 12 | 9 |
| Base Occurrence C | 6 | 17 | 13 | 0 | 0 | 0 | 1 | 0 | 1 | 2 | 2 | 4 | 4 | 39 | 23 | 12 | 21 | 14 |
| Base Occurrence G | 12 | 5 | 4 | 50 | 48 | 0 | 6 | 49 | 43 | 33 | 11 | 28 | 6 | 2 | 10 | 15 | 9 | 4 |
| Base Occurrence T | 14 | 19 | 32 | 0 | 1 | 0 | 41 | 1 | 1 | 7 | 10 | 6 | 11 | 6 | 11 | 11 | 8 | 17 |
| SAAB consensus | N | Y | T | G | G | A | T | G | G | G | R | R | A | C | Y | N | C | Y |
| Eukaryotic consensus | S | Y | N | N | G | R | T | G | G | G | N | G | A | C | Y | N | Y | N |

FIGURE 2: Sequences of box C promoter elements within a synthetic internal control region selected by TFIIIA. The sequences of the noncoding strands of selected ICR templates are shown, along with the frequencies of base occurrence at each randomized position. Single-letter codes in the consensus sequences are based on IUB nomenclature (R = A or G; Y = C or T; S = G or C). The eukaryotic 5S rRNA consensus gene sequence was taken from Erdmann et al. (1985).

appearance of wild-type base pairs in the box C selected sequences is more frequent at the 5′ end (80–87) than at the 3′ end (88–92). This pattern may result from differences in DNA binding affinity of the first three zinc fingers of TFIIIA (Del Rio et al., 1993). Base pairs flanking this recognition sequence (from 78 to 79 and from 93 to 95) remain relatively unselected in the ICR template pool (Figure 2). These positions, however, do exhibit a minor degree of wild-type base pair identity within the selected DNA population and may influence promoter recognition by TFIIIA.

Results of the SAAB assay may also reveal a significant contribution of DNA conformation to TFIIIA binding. This is suggested by the apparent intolerance of many base pair substitutions within the box C element (positions 79, 80, 83, 84, 92, 94, and 95) that result in, at most, minor reductions in TFIIIA binding affinity in the nitrocellulose filter binding

assay. In this respect, the SAAB assay appears to identify the "sequence context" necessary for high-affinity TFIIIA binding in addition to possible individual base pair contacts.

## DISCUSSION

Assembly of an active RNA polymerase III transcription complex on the 5S gene requires recognition of the internal control region by TFIIIA. Three promoter elements are required for the production of a functional transcription complex: the box A, intermediate, and box C elements (Sakonju et al., 1980; Bogenhagen et al., 1980; Pieler et al., 1987). TFIIIA makes sequence-specific contacts within this internal control region (ICR) at three locations: box A (57–62), intermediate (67–70), and box C (78–86) (You et al., 1991). Progressive 5′ deletions that remove the box A promoter element result in a gradual loss of transcription efficiency and corresponding TFIIIA binding ability (Sakonju et al., 1980). In contrast, the reduction in transcriptional competency and TFIIIA binding is more severe for 3′ deletions that enter the box C element (Bogenhagen et al., 1980). Clustered substitution mutations within the 5S RNA gene confirm the importance of the box C element in TFIIIA binding (You et al., 1991). The simultaneous mutation of base pairs 78–86 results in a 100-fold reduction in transcription factor association. The present study further defines TFIIIA contacts within the latter two promoter regions by using an extensive set of point mutants of the 5S RNA gene and a quantitative nitrocellulose filter binding assay for measuring TFIIIA binding affinity. In addition, box C sequences recognized by TFIIIA were isolated and characterized using a SAAB analysis.

The effects of point mutations in the 5S RNA gene on TFIIIA binding affinity do not differentiate between contacts made to bases located on the coding and noncoding DNA strands. A reduction in complex stability due to a single base pair substitution identifies that position but not the individual base as being important for TFIIIA association. This experimental limitation must be kept in mind when interpreting the binding data. However, evidence for the involvement of bases on the noncoding strand in TFIIIA binding to the internal promoter has been elucidated using a variety of methods (Sakonju & Brown, 1982; Lee et al., 1991). It is possible that TFIIIA association with the 5S RNA gene relies predominantly on an extensive set of hydrogen bond contacts to bases on this DNA strand.

TFIIIA interacts with two base pair positions (GC70 and GC71) within the intermediate promoter element (Table 1). Although these contacts have been identified previously (Sakonju & Brown, 1982; Fairall et al., 1986), our data indicate that their contributions to TFIIIA–5S gene complex stability are not equivalent. Mutation of GC70 is more deleterious to TFIIIA binding affinity than is substitution of the adjacent GC71 base pair. Recent studies suggest that zinc finger 5 contacts this base pair doublet in the center of the ICR (Hayes & Clemens, 1992; Clemens et al., 1992).

The apparent discrepancy between results obtained using chemical modification and substitution mutagenesis of the intermediate element may arise from inherent differences between these experimental procedures. Methylation of guanine bases within the ICR could negatively influence interactions between TFIIIA and base contacts adjacent to the modified guanine. The magnitude of such a "neighboring" effect would depend on the energetic contribution provided by the adjacent base contact to the overall free energy of binding and the degree of steric disruption resulting from the introduction of a methyl group in the major groove of the DNA. Thus, studies based on chemical modification of DNA may not allow for an accurate measure of the energetic contribution of individual guanine base contacts within a GC-rich binding site such as the ICR of the 5S RNA gene. Single base pair substitution analysis could provide an advantage in that DNA structure is not likely to be significantly perturbed. In this respect, GC71 is shown to provide a small contribution to TFIIIA interaction with the intermediate element. Methylation of the guanine base at position 71 may abrogate TFIIIA binding to GC70 as well as to GC71, resulting in a greater instability in the TFIIIA–5S RNA gene complex.

DNase I protection by TFIIIA encompasses residues 46–95 of the 5S RNA gene (Engelke et al., 1980; Sakonju & Brown, 1982). Functional studies using 5S DNA point mutants identified the ICR as including base pairs 50–97 (Pieler et al., 1985a,b). However, an investigation of the effects of point mutants centered around base pair 90 suggested that the box C promoter element of the ICR extends only to CG91 (McConkey & Bogenhagen, 1987). TFIIIA contacts within box C have also been defined by methylation and ethylation interference. These data showed that guanine $N^7$ and backbone phosphate contacts extend from TA80 to CG91 (Sakonju & Brown, 1982). This identification of the box C element boundaries is supported by the present point mutant analysis. Base pairs that influence the thermodynamic stability of complex formation, by either direct contact or conformational restrictions, extend from position 80 to 91. Base pairs within this region are also responsible for the high-affinity interaction of a truncated polypeptide containing only the first three zinc fingers of TFIIIA (Liao et al., 1992). SAAB analysis indicates that base identity at position 79 and from position 92 to 94 may influence promoter recognition by TFIIIA, while the mutational data indicate that these base pairs do not make a significant contribution to the free energy of protein binding. Weaker interactions within the box A and intermediate promoter elements may serve to correctly align the bound transcription factor on the internal promoter, producing a biologically functional complex. Oocyte and somatic forms of the 5S RNA gene do exhibit base differences at positions 53, 55, and 56 which lead to significant differences in TFIIIA–5S DNA complex structure and developmental regulation of transcription (Wormington et al., 1981; Sakonju & Brown, 1982; Xing & Worcel, 1989).

TFIIIA interacts with the 5S RNA gene with an apparent association constant ($K_a$) of $1.9 \times 10^9$ $M^{-1}$ and a Gibbs free energy ($\Delta G_c$) of −12.53 kcal/mol (Romaniuk, 1990). In contrast, nonspecific DNA binding by TFIIIA has a $K_a$ of less than $3.3 \times 10^7$ $M^{-1}$ and a $\Delta G_c$ of less than −10.15 kcal/mol. The difference in the free energy of binding ($\Delta\Delta G_c$) of TFIIIA with the internal control region compared to binding with nonspecific DNA is −2.38 kcal/mol. A major contribution to this specific DNA binding energy can be attributed to interactions between TFIIIA and base sequence from position 81 to 91 within the box C promoter element (Table 3). A comparison of the change in free energy of TFIIIA association with clustered and point substitution mutants of the 5S RNA gene was performed to determine the contribution of base sequence and DNA conformation to complex formation (Table 3). Changes in $\Delta G_c$ correlate well between the two types of mutants in the intermediate element, suggesting that individual base sequence positions contribute independently to the binding energy. However, there are moderate differences in $\Delta\Delta G_c$ values for clustered mutants and point mutants within the box C element, indicating that individual base contributions

Table 3: Comparison of the Change in the Free Energy of TFIIIA Association[a] with Clustered and Point Mutants of the 5S RNA Gene

| clustered mutants[b] | | point mutants | |
|---|---|---|---|
| base position | $\Delta\Delta G_c$ | base position | $\Delta\Delta G_c$ |
| | | Intermediate Element | |
| | | 67 | −0.04 ± 0.0003 |
| | | 68 | −0.03 ± 0.0002 |
| | | 69 | 0.05 ± 0.0001 |
| | | 70 | 0.81 ± 0.007 |
| 67–70 | 0.65 ± 0.005 | sum | 0.79 ± 0.007 |
| | | 71 | 0.18 ± 0.001 |
| | | 72 | ND[c] |
| 71–72 | −0.16 ± 0.001 | sum | ND |
| | | Box C Element | |
| | | 78 | −0.12 ± 0.0004 |
| | | 79 | −0.05 ± 0.0004 |
| | | 80 | NS[d] |
| | | 81 | 1.2 ± 0.015 |
| 78–81 | 1.65 ± 0.03 | sum | 1.03 ± 0.015 |
| | | 82 | 0.36 ± 0.003 |
| | | 83 | NS |
| | | 84 | 0.12 ± 0.001 |
| | | 85 | 0.86 ± 0.033 |
| | | 86 | 0.62 ± 0.013 |
| 82–86 | 2.7 ± 0.09 | sum | 1.96 ± 0.033 |
| | | 87 | 0.42 ± 0.003 |
| | | 88 | NS |
| | | 89 | ND |
| | | 90 | 0.54 ± 0.001 |
| 87–90 | 0.26 ± 0.003 | sum | 0.96 ± 0.003 |
| | | 91 | 0.84 ± 0.01 |
| | | 92 | −0.14 ± 0.001 |
| | | 93 | −0.03 ± 0.0001 |
| | | 94 | −0.07 ± 0.003 |
| 91–94 | 0.29 ± 0.006 | sum | 0.6 ± 0.01 |

[a] Determined as the difference in free energies of TFIIIA binding by the mutant nucleic acid and the wild-type nucleic acid. Values are averages of three or more independent determinations. [b] Relative binding affinities for the clustered substitution mutants were reported in You et al. (1991). [c] ND, not determined. [d] NS, base pair not substituted in the clustered mutation; $\Delta\Delta G_c = 0$.

to the free energy of binding are dependent, to some extent, on the sequence context presented to TFIIIA. A recognition process that includes contributions from DNA context could be important in discrimination by a multi-zinc finger protein of the correct promoter sequence from non specific DNA sites that may contain a few correctly positioned contact bases.

Chemical modification, photo-cross-linking, and hydroxy-radical footprinting studies have provided strong evidence for the close association of TFIIIA with the noncoding strand of the ICR (Sakonju & Brown, 1982; Lee et al., 1991; Churchill et al., 1990). Many of the GC base pairs within the 5S gene promoter that make a significant contribution to TFIIIA binding affinity correspond to positions at which methylation of the guanine base on the noncoding DNA strand reduces protein binding (positions 81, 82, 85, 86, 87, and 89) (Figure 3). Reduced TFIIIA binding to the methylated 5S RNA gene indicates a prominent role for the guanine $N^7$ atom in protein interaction. Strong selection, in the SAAB analysis, of guanine residues situated on the noncoding DNA strand at positions 81, 82, 85, 86, 87, and 89 also supports a role for guanine-specific atoms in promoter recognition and binding by TFIIIA. The pattern of hydrogen bond contacts to individual guanine residues may vary significantly along the box C element, as suggested by the differential effects on TFIIIA binding affinity of point mutations at these critical GC base pairs. Selection of the wild-type AT base pair sequence at other positions within the box C region (80, 83, and 84) suggests that they also contribute to the energy of TFIIIA binding, by providing either direct contacts or an optimal DNA conformation.

Contradictory evidence exists that the 5S RNA gene adopts an A-form DNA conformation similar to 5S RNA (McCall et al., 1986; Rhodes & Klug, 1986) or a B-form conformation (Gottesfeld et al., 1987; Aboul-ela et al., 1988; Hayes et al., 1990) similar to mixed-sequence DNA. However, the 5S RNA gene may be structurally heterogeneous, with a conformation intermediate to both classical DNA structures (Fairall et al., 1989; Huber et al., 1991a,b). It has been suggested that this structural heterogeneity results from the periodic positioning of GC-rich sequence clusters along the ICR (Figure 3) (Huber et al., 1991a,b). Such conformational heterogeneity could provide additional information for the recognition and binding of individual zinc fingers of TFIIIA to the 5S RNA gene. A sequence-dependent local DNA conformation required for zinc finger binding may explain selection of native 5S DNA base sequence within the region 79–94 at positions that do not provide energetically significant contributions to TFIIIA binding. AT base pairs selected at positions 79, 80, and 84, as well as CG base pairs at 92 and 94, could help form the local conformation necessary for correct TFIIIA zinc finger positioning in the major groove (Figure 2). Pyrimidines are also found at these latter positions on the noncoding strand in the 5S DNA consensus sequence that includes the 5S RNA genes from highly divergent organisms, such as *Drosophila* and man (Erdmann et al., 1985). Thus, specific DNA conformation, as well as direct base-specific contacts, may help position individual zinc fingers within the box C binding site and align TFIIIA correctly along the ICR.

X-ray crystallographic structures of the DNA binding domains of *zif*268 (Pavletich & Pabo, 1991), *GLI* (Pavletich & Pabo, 1993), and *Tramtrack* (*TTK*) (Fairall et al., 1993) complexed with their respective target DNA sites have been determined. Comparison of individual zinc finger interactions with DNA suggests that no simple rules for determination of zinc finger–DNA contact patterns exist. However, from these few examples, structural parameters of both DNA and zinc finger proteins can be identified that contribute to binding site specificity and protein affinity. DNA subsites contacted by individual zinc fingers range from three to five bases and may be overlapped (*TTK*), abutted (*zif*268), or spaced apart (*GLI*). A single zinc finger can interact with bases on one (*zif*268) or both (*GLI*) DNA strands within a subsite. Interestingly, the DNA conformation can play a major role in binding site specificity by adopting a unique structure that provides critical base and phosphate backbone contacts (Pavletich & Pabo, 1993; Fairall et al., 1993). Amino acid residues at positions −1, 2, 3, 5, and 6, relative to the start of the $\alpha$-helical region within a zinc finger, can be used to contact DNA bases. Differences in the amino acid sequences of zinc fingers result in a variety of base contact patterns that may be difficult to predict through modeling studies. The pattern of phosphate contacts made by each finger can exhibit even greater variability. This allows zinc fingers to be functionally distinct, with individual fingers making predominantly phosphate or base contacts (Pavletich & Pabo, 1993). Thus, although the overall docking of zinc fingers onto DNA appears similar, individual differences in DNA subsite conformation and zinc finger amino acid sequence can result in marked differences in DNA binding specificity. The architecture of the TFIIIA–box C interaction may also exhibit unique characteristics, as suggested by a long contact region and by footprinting studies with TFIIIA broken finger mutants (Del Rio et al., 1993). Individual TFIIIA finger recognition subsites

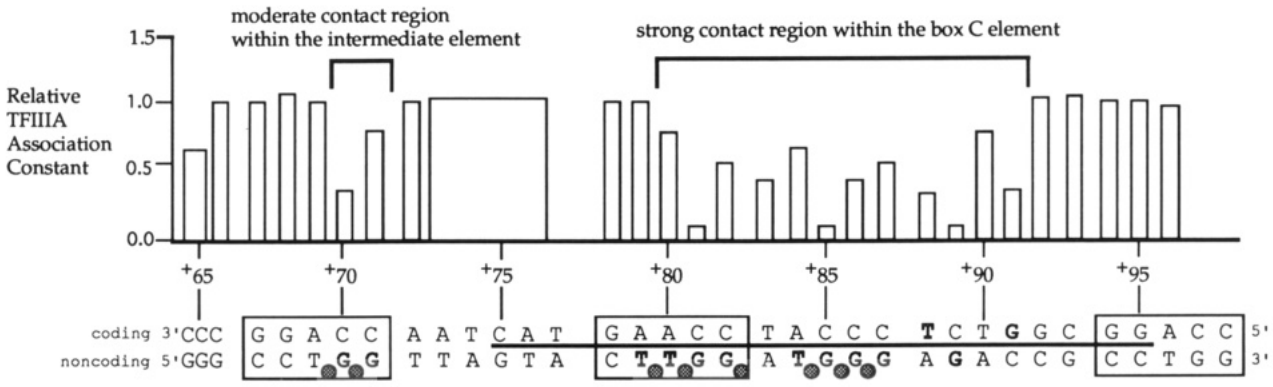



FIGURE 3: Proposed alignment of the N-terminal TFIIIA zinc fingers on the intermediate and box C elements of the 5S RNA gene promoter. Sequences of the coding and noncoding strands from residue 65 to 98 of the 5S RNA gene are shown. The effects of point mutations on TFIIIA binding affinity were averaged for each base pair and are indicated above the DNA sequence. Results obtained using clustered substitution mutations within this region of the ICR are also shown (You et al., 1991). Guanine bases that interfere with TFIIIA binding when methylated are indicated by boldface letters, and DNA backbone phosphates that contact TFIIIA are marked by stippled circles (Sakonju & Brown, 1982). The sequence protected by an N-terminal peptide of TFIIIA containing the first three fingers is underlined (Liao et al., 1992). Thymine residues that closely approach TFIIIA are shown in outlined letters (Lee et al., 1991). Repeated GC-box sequences within this promoter region are boxed (Huber et al., 1991a).

within the box C element could include spacing segments and/or up to five base pair positions.

Truncated mutants of TFIIIA suggest that binding of the N-terminal zinc fingers to the box C element provides a major contribution to the free energy of complex formation (Vrana et al., 1988; Liao et al., 1992; Theunissen et al., 1992). Interactions between the C-terminal zinc fingers and proximal promoter elements further contribute to the binding energy (Clemens et al., 1992; Churchill et al., 1990). This simple model does not take into account possible complex thermodynamic interactions between zinc finger interaction sites that could make it impossible to interpret the energetics of binding by any particular zinc finger considered in isolation. Nonetheless, it appears that the N-terminal zinc fingers of TFIIIA provide nonequivalent contributions to the binding energy. Finger 3 is involved in strong contacts to the box C element, and its removal (Vrana et al., 1988; Liao et al., 1992; Theunissen et al., 1992) or disruption (Del Rio et al., 1993) severely reduces TFIIIA–5S RNA gene binding activity. Finger 2 provides a moderate contribution to the free energy of binding and interacts with the center of the box C element (Del Rio et al., 1993). Deletion of the N-terminal region or finger 1 has little effect on TFIIIA binding affinity. These apparent energetic differences in zinc finger association can be observed in the nonequivalent selection of wild-type sequence along the box C element (Figure 2). Strong selection occurs from 80 to 87, perhaps corresponding to the region of finger 2 and 3 interaction. Weaker selection from 88 to 94 may involve finger 1and amino-terminal nonfinger residues. It is interesting to note that three base pair positions within the box C region (81, 85, and 89) provide the greatest individual contribution to the energy of binding; it is possible that these mutations identify three separate subsites for finger interaction. These positions were also identified as being energetically important in a missing nucleoside analysis of the TFIIIA–5S RNA gene complex (Hayes & Tullius, 1992).

In summary, TFIIIA interacts strongly with the box C element of the 5S gene promoter through multiple contacts to base pairs extending from position 80 to 91. Base pairs flanking this region, at 79 and 92–94, also influence TFIIIA–5S gene association. It is likely that the N-terminal zinc fingers of TFIIIA are involved in this energetically important interaction. Local DNA conformation and critical base contacts are important in promoter recognition by TFIIIA and may function in the alignment of individual zinc fingers along the ICR.

## ACKNOWLEDGMENT

We thank Roberta Hmiel and Shaoyi Liao for identification of the 5S RNA gene mutants used in this study and Dr. Colleen Nelson for helpful discussions and review of the manuscript.

## REFERENCES

Aboul-ela, F., Varani, G., Walker, G. T., & Tinoco, I. (1988) *Nucleic Acids Res. 16*, 3559–3572.
Berg, J. M. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 99–102.
Blackwell, T. K., & Weintraub, H. (1990) *Science 250*, 1104–1110.
Bogenhagen, D. F., Sakonju, S., & Brown, D. D. (1980) *Cell 19*, 27–35.
Bogenhagen, D. F., Wormington, W. M., & Brown, D. D. (1982) *Cell 28*, 413–421.
Bradford, M. M. (1976) *Anal. Biochem. 72*, 248–254.
Brown, R. S., Sander, C., & Argos, P. (1985) *FEBS Lett. 186*, 271–274.
Christensen, J. H., Hansen, P. K., Lillelund, O., & Thøgersen, H. C. (1991) *FEBS Lett. 281*, 181–184.
Churchill, M. E. A., Tullius, T. D., & Klug, A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 5528–5532.
Ciliberto, G., Raugei, G., Constanzo, F., Dente, L., & Cortese, R. (1983) *Cell 32*, 725–733.
Clemens, K. R., Liao, X., Wolf, V., Wright, P. E., & Gottesfeld, J. M. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*,10822–10826.
Del Rio, S., & Setzer, D. R. (1991) *Nucleic Acids Res. 19*, 6197–6203.
Del Rio, S., Menezes, S. R., & Setzer, D. R. (1993) *J. Mol. Biol. 233*, 567–579.
Diakun, G. P., Fairall, L., & Klug, A. (1986) *Nature 324*, 698–699.
Engelke, D. R., Ng, S.-Y., Shastry, B. S., & Roeder, R. G. (1980) *Cell 19*, 717–728.
Erdmann, V. A., Wolters, J., Huysmans, D., & DeWachter, R. (1985) *Nucleic Acids Res. 13*, 103–153.
Fairall, L., Rhodes, D., & Klug, A. (1986) *J. Mol. Biol. 192*, 577–591.
Fairall, L., Martin, S., & Rhodes, D. (1989) *EMBO J. 8*, 1809–1817.
Fairall, L., Schwabe, J. W. R., Capman, L., Finch, J. T., & Rhodes, D. (1993) *Nature 366*, 483–487.
Ginsberg, A. M., King, B. O., & Roeder, R. G. (1984) *Cell 39*, 479–489.

Gottesfeld, J. M., Blanco, J., & Tennant, L. L. (1987) *Nature 329*, 460–462.
Hayes, J. J., & Clemens, K. R. (1992) *Biochemistry 31*, 11600–11605.
Hayes, J. J., & Tullius, T. D. (1992) *J. Mol. Biol. 227*, 407–417.
Hayes, J. J., Tullius, T. D., & Wolffe, A. P. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*,7405–7409.
Huber, P., Blobe, G. C., & Hartman, K. M. (1991a) *J. Biol. Chem. 266*, 3278–3286.
Huber, P. W., Morii, T., Mei, H.-Y., & Barton, J. K. (1991b) *Proc. Natl. Acad. Sci. U.S.A. 88*, 10801–10805.
Lasser, A. B., Martin, P. L., & Roeder, R. G. (1983) *Science 222*, 740–748.
Lee, D. K., Evans, R. K., Blanco, J., Gottesfeld, J. M., & Johnson, J. D. (1991) *J. Biol. Chem. 266*, 16478–16484.
Liao, X., Clemens, K. R., Tennant, L., Wright, P. E., & Gottesfeld, J. M. (1992) *J. Mol. Biol. 223*, 857–871.
McCall, M., Brown, T., Hunter, W. N., & Kennard, O. (1986) *Nature 322*, 661–664.
McConkey, G. A., & Bogenhagen, D. F. (1987) *Mol. Cell. Biol. 7*, 486–494.
Miller, J., McLachlan, A. D., & Klug, A. (1985) *EMBO J. 4*, 1609–1614.
Pavletich, N. P., & Pabo, C. O. (1991) *Science 252*, 809–817.
Pavletich, N. P., & Pabo, C. O. (1993) *Science 261*, 1701–1707.
Pelham, H. R. B., & Brown, D. D. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 4170–4174.
Picard, B., le Maire, M., Wegnez, M., & Denis, H. (1980) *Eur. J. Biochem. 109*, 359–368.
Pieler, T., Appel, B., Oei, S.-L., Mentzel, H., & Erdmann, V. A. (1985a) *EMBO J. 4*, 1847–1853.
Pieler, T., Oei, S-L., Hamm, J., Engelke, U., & Erdmann, V. A. (1985b) *EMBO J. 4*, 3751–3756.
Pieler, T., Hamm, J., & Roeder, R. G. (1987) *Cell 48*, 91–100.
Rhodes, D., & Klug, A. (1986) *Cell 46*, 123–132.
Romaniuk, P. J. (1985) *Nucleic Acids Res. 13*, 5369–5387.
Romaniuk, P. J. (1990) *J. Biol. Chem. 265*, 17593–17600.
Sakonju, S., & Brown, D. D. (1981) *Cell 23*, 665–669.
Sakonju, S., & Brown, D. D. (1982) *Cell 31*, 395–405.
Sakonju, S., Bogenhagen, D. F., & Brown, D. D. (1980) *Cell 19*, 13–25.
Setzer, D. R., Hmiel, R. M., & Liao, S. (1990) *Nucleic Acids Res. 18*, 4175–4178.
Shastry, B. S., Ng, S.-Y., & Roeder, R. G. (1982) *J. Biol. Chem. 257*, 12979–12986.
Smith, D. R., Jackson, I. J., & Brown, D. D. (1984) *Cell 37*, 645–652.
Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60–88.
Theunissen, O., Rudt, F., Guddat, U., Mentzel, H., & Pieler, T. (1992) *Cell 71*, 679–690.
Vrana, K. E., Churchill, M. E. A., Tullius, T. D., & Brown, D. D. (1988) *Mol. Cell. Biol. 8*, 1684–1696.
Wormington, W. M., Bogenhagen, D. F., Jordan, E., & Brown, D. D. (1981) *Cell 24*, 809–817.
Xing, Y. Y., & Worcel, A. (1989) *Mol. Cell. Biol. 9*, 499–514.
You, Q., Veldhoen, N., Baudin, F., & Romaniuk, P. J. (1991) *Biochemistry 30*, 2495–2500.