Adenines at –11, –9 and –8 play a key role in the binding of *Bacillus subtilis* $E\sigma^A$ RNA polymerase to –10 region single-stranded DNA

Jian Qiu¹ and John D. Helmann^{1,2,*}

¹Field of Biochemistry, Molecular and Cell Biology and ²Department of Microbiology, Cornell University, Ithaca, NY 14853, USA

Received September 9, 1999; Revised and Accepted October 19, 1999

ABSTRACT

The σ subunit of RNA polymerase interacts with the promoter DNA in at least two regions: the -35 and the -10 consensus elements. The latter contacts are involved both in recognition and in melting of the promoter DNA to form the transcriptionally-competent open complex. RNA polymerase holoenzyme, but neither core nor σ alone, binds with high selectivity to single-stranded DNA (ssDNA) containing the nontemplate -10 consensus sequence. We have used equilibrium competition to assess the specificity of holoenzyme binding to a 19 base oligonucleotide containing a -10 consensus element, TATAAT. Analysis of all 18 possible single point mutations in the -10 consensus sequence reveals that binding by *Bacillus subtilis* $E\sigma^A$ holoenzyme depends critically upon adenine at position -11 and, unexpectedly, is strongly affected by substitutions of the poorly conserved adenines at -9 and -8. Similarly, ssDNA binding by Escherichia coli Eo⁷⁰ holoenzyme is most strongly affected by substitutions of adenines within the -10 region consensus. The critical role of -11A in binding ssDNA supports a key role for this base in the nucleation of DNA melting. A novel role for -9A and -8A is proposed in the context of recent models of promoter melting.

INTRODUCTION

The selection of transcription start sites is determined by the sequence-specific recognition of promoters by RNA polymerase (RNAP) and depends critically on the associated σ subunit (1–3). In most bacteria, a single primary σ factor controls the majority of transcription in growing cells. In *Escherichia coli*, the primary σ is σ^{70} while the corresponding factor in *Bacillus subtilis* is designated σ^A . RNAP holoenzyme containing either of these proteins preferentially recognizes promoters with consensus elements of TTGaca (–35) and TAtaaT (–10), where the positions indicated in uppercase are

the most highly conserved and generally considered to be the most important for determining promoter strength (4,5).

RNAP interacts with promoter DNA over a large region of ~80 bp extending from near -60 to +20 relative to the transcription start point. The most critical determinants for promoter recognition are generally the consensus hexamers located near -35 and -10. However, the role of the -35 region is dispensible and can be functionally replaced by either activator protein—RNAP contacts or by the presence of an additional conserved sequence near -15 in 'extended -10' promoters (4,6,7). Additional important contacts are often present in the upstream promoter (UP) region between -65 and -40, where the C-terminal domain of the α subunit can form either sequence-dependent or sequence-independent contacts (8–10). Together, these various contacts determine the overall affinity of RNAP for the promoter region.

Promoter strength reflects not only the initial affinity of RNAP for the promoter region, but the facility with which the bound enzyme can melt DNA, initiate an RNA chain, and escape from abortive cycling and/or promoter proximal pauses to form an elongation complex (4). In general, strong promoter sites often match closely to the key consensus elements, but too close a match can ultimately reduce promoter strength by impeding the ability of RNAP to escape from its tight contacts with the promoter DNA (11).

We are investigating the role of the σ subunit in -10 region recognition and DNA melting, with an emphasis on the contacts that occur between conserved region 2 of σ and the -10 region consensus sequence (12–14). The structure of this region in a fragment of *E.coli* σ^{70} , revealed by X-ray crystallography (15), provides a framework for thinking about these interactions. Most of the amino acids implicated in promoter recognition (conserved σ region 2.4) or melting (region 2.3) project from a single α helix (helix 14). Genetic suppression analysis suggests that one or more amino acids from this helix are responsible for specifying the conserved T at position -12 (16), but suppression studies did not reveal allele-specific interactions with other positions in the -10 element and the amino acids that specify these positions are not yet clear.

We have previously implicated four conserved aromatic amino acids (*E.coli* σ^{70} Y425, Y430, W433 and W434) exposed on the surface of helix 14 in the stabilization of single-stranded DNA (ssDNA) formed during promoter melting (12–14). As one

*To whom correspondence should be addressed at: Department of Microbiology, Cornell University, Ithaca, NY 14853-8101, USA. Tel: +1 607 255 6570; Fax: +1 607 255 3904; Email: jdh9@cornell.edu

approach to study the interactions between RNAP and ssDNA, we have taken advantage of the observation that RNAP holoenzyme binds with high affinity and selectivity to ssDNA containing a -10 consensus element and this interaction likely involves region 2 of σ factor (17). This model is supported by the observations that sequence selective ssDNA binding requires σ , truncated σ factors containing region 2 suffice for this interaction, and bound DNA is crosslinked to σ by UV irradiation (17). In addition, point mutations in region 2 that affect promoter melting alter the affinity and/or crosslinking of ssDNA to σ (14) and ssDNA effectively quenches the fluorescence of amino acid analogs incorporated in place of the two Trp residues implicated in promoter melting (18). It is likely that the interactions between ssDNA and RNAP as measured in this assay closely mimic those found in open complexes, an assumption supported by studies using an altered specificity σ factor and a corresponding mutant -10 element (17).

The details of the contacts between σ region 2 and ssDNA in the open complex, and in the oligonucleotide model system, are not yet clear, but the high selectivity observed suggests that some or all of -10 region promoter selectivity may be conferred by interactions between ssDNA, rather than double-stranded DNA (dsDNA) as generally assumed. To further explore the sequence selectivity of this binding reaction, we have tested the effects of all 18 possible point mutations on binding between ssDNA oligonucleotides and RNAP holoenzyme. The results with *B.subtilis* RNAP indicate that the three adenine bases are the most critical determinants of ssDNA selectivity.

MATERIALS AND METHODS

Materials

Bacillus subtilis RNAP core enzyme and σ^A preparations have been described previously (12,19). *Escherichia coli* RNAP σ^{70} -saturated holoenzyme was kindly provided by C. Turnbough (University of Alabama, Birmingham, AL). This enzyme was purified by the Burgess and Jendrisak method (20) followed by an FPLC MonoQ chromatography step (21) that separates core from holoenzyme.

The oligonucleotides used in these studies include the previously described consensus oligonucleotide, (C) 5'-ATT-GGG<u>TATAAT</u>TGACTCA-3' (17), and variants that differ from consensus by single base substitutions at the underlined positions, extending from -12T (5'-end) to -7T (3'-end).

DNA binding assays

Electrophoretic mobility shift assays (EMSA) were performed by pre-mixing ³²P-labeled oligonucleotide C with the competing oligonucleotide prior to addition of RNAP (14). Binding reactions (20 µl) contained 2 nM labeled oligonucleotide C, 200 nM *B.subtilis* core RNAP, 3 µM σ^A , and competing oligonucleotide at the indicated ratio relative to 200 nM RNAP. Reactions containing *E.coli* σ^{70} holoenzyme contained a final concentration of 880 nM RNAP. RNAP concentrations have not been corrected for percent active molecules and refer to total protein. Thus, the molar excess relative to active RNAP is even higher.

Reactions were incubated at room temperature for 30 min in binding buffer (20 mM Tris–HCl, pH 8.0, 10 mM NaCl, 10 mM MgCl₂, 0.1 mM EDTA, 1 mM dithiothreitol, 5% v/v

glycerol and 100 μ g/ml BSA), 2 μ l of loading buffer (0.1% bromophenol blue, 0.1% xylene cyanol FF and 50% glycerol) was added and the reactions fractionated by 4% PAGE run in 1×TAE buffer at 180 V for 1 h at room temperature. Gels were pre-electrophoresed for 60 min prior to sample addition. After electrophoresis, gels were dried and visualized by autoradiography. To quantify the amount of DNA bound, a Phosphor-Imager (Molecular Dynamics) was used together with Image-Quant data analysis software.

RESULTS

Sequence selectivity of *B.subtilis* σ^A holoenzyme

To investigate the sequence selectivity of ssDNA binding by RNAP holoenzyme, we pre-mix the radiolabeled consensuscontaining oligonucleotide, C, with various concentrations of competing oligonucleotide. As reported previously (14,17), the binding of C to RNAP is σ -dependent: a slower mobility complex, as visualized by native gel electrophoresis, forms with holoenzyme, but only inefficiently with core alone (data not shown). As a control, a self-competition reaction is included. As expected, the dilution of labeled C with unlabeled C oligonucleotide leads to a decrease in the amount of radiolabel in the shifted complex. Previous studies have demonstrated that this binding is specific to the non-template strand: little or no competition is observed with the template strand, an anticonsensus oligonucleotide, or a DNA duplex containing the -10 consensus (14,17).

To determine which of the bases within the -10 consensus element are important for selective binding we measured the ability of RNAP to bind to labeled C oligonucleotide in the presence of excess mutant oligonucleotides. Representative results for selected mutations at the -12T and -11A positions are shown for both the *B.subtilis* σ^A (Fig. 1A) and *E.coli* σ^{70} (Fig. 1B) holoenzymes.

For *B.subtilis* σ^A holoenzyme, we find that unlabeled oligonucleotides containing alterations at the -12T position compete for the ssDNA binding site nearly as effectively as the consensus oligonucleotide (containing T at this position). Thus, -12T does not appear to be a critical determinant for ssDNA recognition. In contrast, mutation of the -11A to either G (Fig. 1A) or T (Fig. 1B) greatly reduces the ability of RNAP to recognize this competing oligonucleotide: even a large molar excess of the competing oligonucleotide does not prevent efficient binding of the labeled C oligonucleotide.

A summary of the competition results for the *B.subtilis* σ^A holoenzyme is presented in Figure 2. Note that a 2.5-fold molar excess of C oligonucleotide decreases binding of the labeled oligonucleotide to ~10% of the control. If all of the holoenzyme was active for ssDNA binding, we would expect a residual binding of ~40%. This suggests that only ~25% of the holoenzyme is active in this assay. Analysis of mutants at all six positions in the consensus sequence reveals that the *B.subtilis* σ^A holoenzyme strongly discriminates against oligonucleotides altered in the –11A position (Fig. 2). The next most important positions include the A residues at –9 and –8. Interestingly, a hierarchy of base preferences is noted at the –8 position, and less so at most other positions. Specifically, the binding preference at this position is TATAAT > TATACT > TATAGT > TATATT. Note that two of the positions that

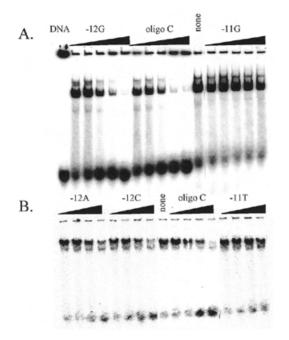


Figure 1. Representative results from EMSA using RNAP holoenzyme, labeled C oligonucleotide and increasing amounts of competitor oligonucleotide. (**A**) *Bacillus subtilis* σ^A holoenzyme. The competing oligonucleotides (containing the indicated base change from consensus) were added in increasing molar excess (0.025-, 0.1-, 0.5-, 2.5- and 10-fold) relative to 200 nM RNAP holoenzyme. (**B**) *Escherichia coli* σ^{70} holoenzyme with oligonucleotides (containing the indicated base change) added in increasing molar excess (0.2-, 1.1-, 4.5- and 23-fold) relative to 880 nM RNAP holoenzyme.

display relatively little effect on the ssDNA binding affinity, -12T and -7T, are (together with -11A) among the most highly conserved positions in bacterial promoters (22,23).

The origins of these sequence preferences are not clear, but we have hypothesized that this region of DNA may be in close contact with aromatic amino acids from region 2.3 (24,25). To determine whether there are allele-specific interactions between the mutant oligonucleotides and σ factors altered in region 2, we individually tested the effects of σ^A mutants Y184A, Y189L, W192A and W193A on the sequence selectivity of the resulting holoenzymes. However, in nearly every case the mutant holoenzymes displayed a selectivity comparable, or even slightly increased, relative to that of the wild-type holoenzyme (data not shown). This suggests that although these residues may contact this region of ssDNA, they do not contribute significantly to sequence selectivity under the conditions of this assay.

Sequence selectivity of *E.coli* **o**70 holoenzyme

To detemine whether the results obtained with the *B.subtilis* holoenzyme were general, we repeated these studies using *E.coli* σ^{70} holoenzyme (Fig. 3). In contrast to the *B.subtilis* holoenzyme, the competition pattern observed for the *E.coli* enzyme was not as dramatic. In part, this reflects the lower activity of this enzyme preparation: greater amounts of total enzyme were required to shift the majority of the labeled probe in the absence of competition. Furthermore, the *E.coli* holoenzyme bound considerable amounts of probe DNA even in the presence of a molar excess of competitor. The reasons for

this comparatively high background of DNA-binding activity are not clear.

In general, the determinants necessary for sequence-selective binding by the *E.coli* holoenyzme appear to be more broadly distributed throughout the -10 consensus element, but the -11A position again emerges as most important: mutations in this base essentially abolish the ability of the mutant oligonucleotides to compete for the ssDNA-binding site (Figs 1B and 3). As noted for the *B.subtilis* holoenzyme, the -12 and -7T positions are again the least important for sequence selective binding. The effects of mutations at -10T, -9A and -8A are intermediate in nature.

DISCUSSION

We have used an equilibrium competition binding assay to assess the ability of RNAP holoenzyme to bind to a -10 region consensus oligonucleotide in the presence of a large excess of mutant oligonucleotides altered in a single base. The results demonstrate that the ssDNA binding site of the *B.subtilis* holoenzyme is exquisitely sensitive to even single base changes. However, there is an imperfect correlation between those positions sensed by holoenzyme in ssDNA, and those positions most important for promoter function (Fig. 4A).

Several lines of evidence indicate that the three -10 region positions most important for promoter function are <u>TAtaaT</u>. Mutations at these positions have the largest effects on promoter strength and compilations of σ^{70} -type promoters from *E.coli* (22) and *B.subtilis* (23) reveal that these positions are the most conserved. In contrast, we have found that the determinants for high affinity ssDNA binding by *B.subtilis* σ^{A} holoenzyme include only one of these three functionally important bases, -11A. A similar overall pattern, albeit much less dramatic, is observed with the *E.coli* holoenzyme.

The *B.subtilis* holoenzyme does not display a strong base preference at position -12T despite the critical role this base plays in promoter function. Thus, we propose that recognition of this position may occur principally in dsDNA. However, sequence-selective recognition of this position in ssDNA has been demonstrated previously: a -12T to C mutation was found to have a modest effect on binding of ssDNA by E.coli holoenzyme and this effect could be reversed by the presence of a corresponding mutation in $\sigma^{70}(17)$. We also observe some sequence selectivity for this position in the ssDNA binding assay when using E.coli holoenyzme (Fig. 3). However, in other studies a substitution at this position had very little effect on the binding of fluorescently-labeled ssDNA oligonucleotides (26). Thus, the modest effects of this position on binding affinity seem to vary depending both on the source of holoenzyme and the precise reaction conditions.

Similarly, our studies do not demonstrate a strong base preference at position -7 for binding ssDNA. This is consistent with studies using 'fork-junction' templates containing an overhanging -10 region non-template strand: the identity of the base at -7 was found to be unimportant for the formation of heparin resistant open complexes (27). In contrast, sequence selectivity at this position was observed in transcription experiments using templates containing various base mismatches at this position (28). It is certainly possible that recognition of -7might precede promoter melting: kinetic studies using UV laser crosslinking have indicated that RNAP contacts DNA to

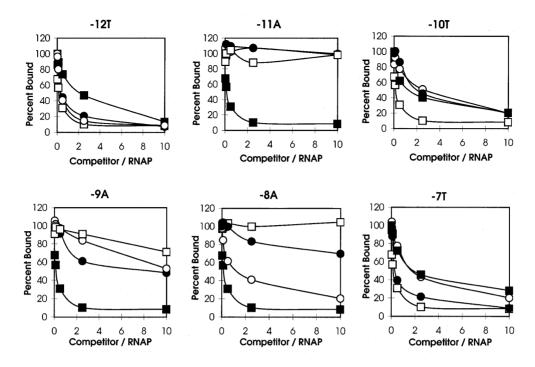


Figure 2. Summary of EMSA data for *B.subtilis* σ^A holoenzyme. Each panel represents the results of binding of RNAP to the labeled C oligonucleotide in the presence of increasing amounts of four competing oligonucleotides (one of which is C). The base at the indicated position in the unlabeled competing oligonucleotide is either T (open square), A (closed square), C (open circle) or G (closed circle). Note that in every case, the most efficient competition is observed with the oligonucleotide containing the consensus base.

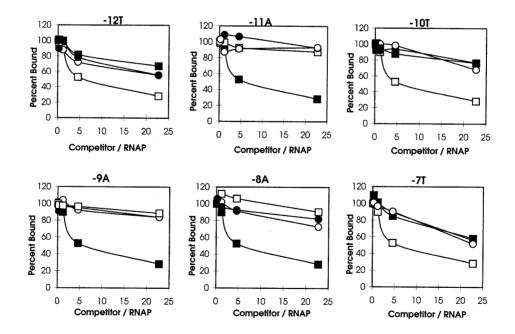


Figure 3. Summary of EMSA data for *E.coli* σ^{70} holoenzyme. Each panel represents the results of binding of RNAP to the labeled C oligonucleotide in the presence of increasing amounts of four competing oligonucleotides (one of which is C). The base at the indicated position in the unlabeled competing oligonucleotide is either T (open square), A (closed square), C (open circle) or G (closed circle). Note that in every case, the most efficient competition is observed with the oligonucleotide containing the consensus base.

at least -3 on the non-template strand prior to strand separation (29).

The critical role of the -11A residue both for promoter function and ssDNA binding is consistent with several other studies (see 25 for review). For example, a single overhanging -11A has been found to suffice for high-affinity binding of 'fork-junction' templates to *E.coli* σ^{70} holoenzyme (27). In experiments using DNA mismatches, it was clearly demonstrated that recognition

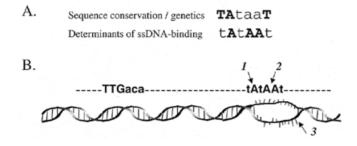


Figure 4. (A) Comparison of base preferences in the -10 consensus element as deduced from promoter sequence conservation and mutational analyses (top) and ssDNA binding assays as reported here (bottom). (B) Structure of a mature open complex (see 4 and 25 for reviews). Formation of the transcription bubble is proposed to involve (1) a critical interaction of the -11A with a σ -dependent, ssDNA-specific binding site on RNAP, (2) an initial expansion in which the non-template strand associates with the ssDNA binding site (a reaction favored by the presence of A at -9 and -8) and (3) the Mg²⁺-dependent isomerization to the mature open complex in which the template strand fills the active site.

of this position requires the non-template A rather than the template strand T (28). This position is unique, among all six positions of the -10 hexamer, in that depurination at this site (loss of the non-template strand adenine) leads to a decrease, rather than an increase, in the rate of open complex formation (30). Together, these data support a model in which recognition of this base occurs concomitant with DNA melting (25).

Our studies revealed a very marked sequence hierarchy for recognition -8A by the *B.subtilis* holoenzyme (A > C > G > T). The modest effects of C substitutions are consistent with the observed base conservation at this position. Compilations of B.subtilis (23) and E.coli (22,31) promoters reveal a base preference of A > C > T > G. This hierarchy is further supported by studies of all possible point mutations in the consensus elements of P22 ant promoter which demonstrate that a C at -8 retains 72% of wild-type activity, while either a G or a T reduces activity to <25% of wild-type (32). Further corroboration for this preference is provided by experiments in which random hexamers were selected for -10 region function in the context of either a consensus -35 element (33) or the malP1 promoter (34). In each case, the selected sequence displayed A and C residues at -8 in a greater fraction of the recovered promoters than either G or T. Based on our ssDNA binding studies we suggest that position -8 is recognized by RNAP subsequent to DNA melting.

The lack of correspondence between those bases most important for promoter function, and those most important for ssDNA binding (Fig. 4A), likely derives from the several differences between these two assays. Promoter function reflects a composite measure of the ability of RNAP to bind to the DNA, form a strand-separated open complex, and initiate and clear the promoter region (4,6). The initial binding reaction almost certainly involves contacts both to ds- and ssDNA regions. In contrast, our assay only measures binding between RNAP and non-template ssDNA. We propose that RNAP interacts first with the -35 and UP-element regions and then the DNA wraps around the enzyme to allow interactions between σ region 2 and the still double-stranded -10 region

(including the conserved -12T and perhaps -7T). Nucleation of melting occurs when the -11A nucleotide undergoes a 'base-flipping' reaction (35) and the transiently exposed adenine establishes sequence-specific interaction with holoenzyme (Fig. 4B, step 1). This step is often rate-limiting and thus can account for the critical role of -11A both for promoter function in vivo and for ssDNA binding in vitro. Based on our binding studies, we propose that the preferred adenine bases at positions -9 and -8 interact with the ssDNA-binding site (step 2) to help mediate the downstream extension of the transcription bubble. Finally, the bases of the template strand associate with the active site of the enzyme in a Mg^{2+} -dependent (4) isomerization step (step 3). At promoters where bubble expansion (steps 2 and 3) is rapid compared to the rate-limiting step (often either binding or nucleation; step 1), changes at -8 and -7 are not predicted to alter the kinetics of the initiation process or to have a dramatic effect on strength.

ACKNOWLEDGEMENTS

We thank Dr Charles Turbough (University of Alabama, Birmingham, AL) and Dr Jeff Roberts (Cornell University, Ithaca, NY) for generously supplying *E.coli* RNAP, and Dr Pieter deHaseth (Case Western Reserve University, Cleveland, OH) for helpful comments on this manuscript. This work was supported by a grant from the National Institutes of Health (GM47446) and by a sub-contract from grant MDA 972-96-K-0003 from the Defense Advance Research Projects Agency (DARPA) awarded to Dr Larry DeLucas (University of Alabama, Birmingham, AL).

REFERENCES

- Helmann, J.D. (1994) In Conaway, R.C. and Conaway, J.W. (eds), *Transcription: Mechanisms and Regulation*. Raven Press, New York, pp. 1–17.
- Gross, C.A., Lonetto, M. and Losick, R. (1992) In McKnight, S.L. and Yamamoto, K.R. (eds), *Transcriptional Regulation*, Vol. 1. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 129–176.
- 3. Helmann, J.D. and Chamberlin, M.J. (1988) *Annu. Rev. Biochem.*, **57**, 839–872.
- 4. deHaseth,P.L., Zupancic,M. and Record,M.T.,Jr (1998) J. Bacteriol., 180, 3019–3025.
- 5. Travers, A.A. (1987) CRC Crit. Rev. Biochem., 22, 181–198.
- 6. Roy, S., Garges, S. and Adhya, S. (1998) J. Biol. Chem., 273, 14059–14062.
- Bown, J., Barne, K., Minchin, S. and Busby, S. (1997) Nucleic Acids Mol. Biol., 11, 41–52.
- Ross, W., Gosink, K.K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K. and Gourse, R.L. (1993) *Science*, 262, 1407–1413.
- 9. Ross, W., Aiyar, S.E., Slomon, J. and Gourse, R.L. (1998) J. Bacteriol., 180, 5375–5383.
- Tang, Y., Murakami, K., Ishihama, A. and deHaseth, P.L. (1996) J. Bacteriol., 178, 6945–6951.
- Ellinger, T., Behnke, D., Bujard, H. and Gralla, J.D. (1994) J. Mol. Biol., 239, 455–465.
- 12. Juang, Y.-L. and Helmann, J.D. (1994) J. Mol. Biol., 235, 1470–1488.
- 13. Rong, J.C. and Helmann, J.D. (1994) J. Bacteriol., 176, 5218–5224.
- Huang, X., Lopez de Saro, F.J. and Helmann, J.D. (1997) Nucleic Acids Res., 25, 2603–2609.
- 15. Malhotra, A., Severinova, E. and Darst, S.A. (1996) Cell, 87, 127-136.
- Waldburger, C., Gardella, T., Wong, R. and Susskind, M.M. (1990) J. Mol. Biol., 215, 267–276.
- 17. Marr, M.T. and Roberts, J.W. (1997) Science, 276, 1258–1260.
- Callaci,S., Heyduk,E. and Heyduk,T. (1998) J. Biol. Chem., 273, 32995–33001.
- 19. Juang, Y.L. and Helmann, J.D. (1994) J. Mol. Biol., 239, 1–14.
- 20. Burgess, R.R. and Jendrisak, J.J. (1975) Biochemistry, 14, 4634-4638.

- Hager, D.A., Jin, D.J. and Burgess, R.R. (1990) Biochemistry, 29, 7890–7894.
- 22. Lisser, S. and Margalit, H. (1993) Nucleic Acids Res., 21, 1507-1516.
- 23. Helmann, J.D. (1995) Nucleic Acids Res., 23, 2351–2360.
- 24. deHaseth,P.L. and Helmann,J.D. (1995) Mol. Microbiol., 16, 817-824.
- 25. Helmann, J.D. and deHaseth, P.L. (1999) Biochemistry, 38, 5959-5967.
- Fedoriw, A.M., Liu, H., Anderson, V.E. and deHaseth, P.L. (1998) Biochemistry, 37, 11971–11979.
- 27. Guo, Y. and Gralla, J.D. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 11655–11660.
- 28. Roberts, C.W. and Roberts, J.W. (1996) Cell, 86, 495-501.

- Buckle, M., Pemberton, I.K., Jacquet, M.A. and Buc, H. (1999) J. Mol. Biol., 285, 955–964.
- 30. Li,X.-Y. and McClure,W.R. (1998) J. Biol. Chem., 273, 23558-23566.
- 31. Hawley, D.K. and McClure, W.R. (1983) *Nucleic Acids Res.*, **11**, 2237–2255.
- Moyle,H., Waldburger,C. and Susskind,M.M. (1991) J. Bacteriol., 173, 1944–1950.
- 33. Oliphant, A.R. and Struhl, K. (1988) Nucleic Acids Res., 16, 7673-7683.
- 34. Danot, O. and Raibaud, O. (1994) J. Mol. Biol., 238, 643-648.
- 35. Roberts, R.J. and Cheng, X. (1998) Annu. Rev. Biochem., 67, 181-198.