Targeted Protein Footprinting: Where Different Transcription Factors Bind to RNA Polymerase†

Stacey L. Traviglia,‡ Saul A. Datwyler,‡ Dalai Yan,§ Akira Ishihama,|| and Claude F. Meares*,‡

Department of Chemistry, University of California, One Shields Avenue, Davis, California 95616, Department of Plant and Microbial Biology and Molecular and Cell Biology, University of California, Berkeley, California 94720, and Department of Molecular Genetics, National Institute of Genetics, Mishima, Shizuoka 411, Japan

Received July 26, 1999; Revised Manuscript Received September 17, 1999

ABSTRACT: Gene transcription is regulated through the interactions of RNA polymerase (RNAP) with transcription factors, such as the bacterial σ proteins. We have devised a new strategy that relies on targeted protein footprinting to make an extensive survey of proximity to the protein surface. This involves attaching cutting reagents randomly to lysine residues on the surface of a protein such as σ. The lysine-labeled σ protein is then used to cleave the polypeptide backbones of the RNAP proteins at exposed residues adjacent to the σ binding site. We used targeted protein footprinting to compare the areas near which σ^{70}, σ^{54}, σ^{38}, σ^{E}, NusA, GreA, and omega bind to the protein subunits of *Escherichia coli* RNAP. The σ proteins and NusA cut sites in similar regions of the two large RNAP subunits, β and β', outlining a common surface. GreA cuts a larger set of sites, whereas omega shows no overlap with the others, cutting only the β' subunit at a unique location.

Many processes important for the development and survival of organisms involve the binding of regulatory proteins to large macromolecular complexes. To understand these molecular events, the questions to be addressed begin with identifying protein−protein or protein−nucleic acid contacts. Chemical cleavage with tethered reagents has proved to be particularly well suited for mapping sites in macromolecular complexes (*1*−*3*). The focus of the present investigation, the RNA polymerase (RNAP) of *Escherichia coli*, illustrates the power of this approach. Core RNAP,[1] a noncovalent complex of four proteins ($\alpha_2\beta\beta'$), binds to one of several different σ proteins to form the holoenzyme ($\alpha_2\beta\beta'\sigma$). Each distinct σ directs RNAP to a characteristic promoter site on DNA. Shortly after RNA synthesis begins, σ is released and other regulatory proteins such as NusA can bind to the transcription complex (*4*).

The binding sites of regulatory proteins on core RNAP are not well understood, primarily due to a lack of experimental tools. We have recently developed an efficient method, targeted protein footprinting, for mapping regions near the binding site of one protein on another (*5*). This involves attaching cutting reagents randomly to lysine residues on a protein's surface and then using this lysine-labeled protein to cleave the polypeptide backbone of the other protein at exposed residues adjacent to its binding site (Figure 1). The cleavage of nucleic acids with lysine-labeled proteins has also been described (*6*, *7*). Because lysines are common residues on the surfaces of most proteins, this new methodology makes it straightforward to label any number of proteins, allowing many different complexes to be studied in parallel. This technique was first used to explore those regions of the σ^{70} binding site on core RNAP that had not been accessed by cutting reagents attached to single-cysteine mutants. Targeted protein footprinting by lysine-labeled σ^{70} provided information equivalent to a large set of single-cysteine mutants. This has led us to explore and compare the interactions of several different proteins with RNAP.

An important feature of targeted protein footprinting is that a macromolecule bound to *any* site on the lysine-labeled protein should be cleaved. This requires that the molecules in a lysine-labeled protein preparation contain cutters located at diverse sites, such that the entire surface of the lysine-labeled protein is collectively reached by cutting reagents (Figure 1). The crystal structures of several proteins reveal that lysines are distributed over their surfaces, with the average lysine ϵ-amino group about 12 Å away from the corresponding group on its nearest neighboring lysine. We modify the native lysine side chains of a chosen protein with the cutting reagent iron (*S*)-1-(*p*-bromoacetamidobenzyl)-ethylenediaminetetraacetate (FeBABE) via a 2-iminothiolane (2IT) spacer. Each tethered 2IT-FeBABE can reach 18 Å, sufficient to extend past the nearest neighboring lysine. As Figure 2 shows, almost complete surveillance of the surface can be achieved for a variety of proteins (*8*−*10*). For cutting caused by diffusible species such as hydroxyl radicals, the reach of each tethered reagent will be even greater.

The technology described here allows for efficient screening of protein−protein interactions and makes it easy to

† This work was supported by Research Grant GM25909 to C.F.M. from the National Institutes of Health.

* Address correspondence to this author. Telephone: 530-752-0936. Fax: 530-752-8938. E-mail: cfmeares@ucdavis.edu.

‡ University of California, Davis.

§ University of California, Berkeley.

|| National Institute of Genetics.

[1] Abbreviations: RNAP, RNA polymerase from *Escherichia coli*; FeBABE, iron (*S*)-1-(*p*-bromoacetamidobenzyl)ethylenediaminetetraacetate; 2IT, 2-iminothiolane; BSA, bovine serum albumin; CNBr, cyanogen bromide; EDTA, ethylenediaminetetraacetate; MOPS, 3-morpholinopropanesulfonate; NTCB, 2-nitro-5-thiocyanobenzoic acid.

10.1021/bi9917232 CCC: $18.00
Published on Web 11/11/1999

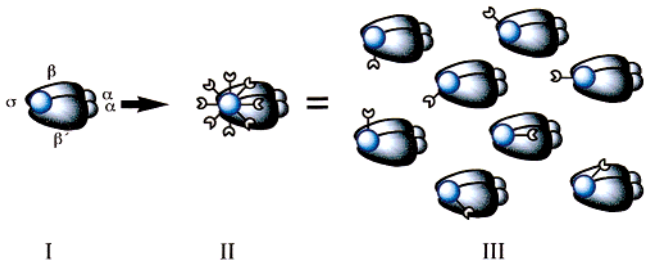

FIGURE 1: General approach for mapping protein−protein interaction sites by targeted protein footprinting, illustrated for RNA polymerase (I). To characterize the interaction of a sigma protein (blue) with core RNA polymerase, a cutting reagent should be placed at many locations on the surface of sigma (II). This is achieved experimentally with a collection of lysine-labeled protein molecules, with each molecule containing one cutting reagent at a different location (III). Since the side chain ϵ-amino groups of lysine residues tend to have similar reactivity, the cutting reagent can be distributed among surface-exposed lysines. A low degree of conjugation ensures that no single protein molecule is likely to have more than 1−2 cutting reagents attached, and modification can be accomplished with only moderate loss of biological activity (*5*). Upon activation, the tethered iron chelate (2IT-FeBABE) cleaves the adjacent segments of proteins and nucleic acids.

assess where multiple factors bind to a common target. We employed this new approach to determine where different proteins bind to the RNAP core enzyme.

MATERIALS AND METHODS

Materials. RNAP and RNAP binding proteins were purified according to literature references (*11*−*18*). The bovine serum albumin (BSA) used for control experiments was purchased from Pierce. The RNAP binding proteins were conjugated with the cutting reagent, FeBABE (*19*), via a 2-iminothiolane (2IT) linker (purchased from Sigma). The cleavage reactions were initiated with ascorbic acid (Fluka; vitamin C, microselect grade) and hydrogen peroxide (J. T. Baker; Ultrex grade). Pure water (18 MΩ cm) was used for all experiments.

2IT-FeBABE Conjugation. The conjugation reactions were performed as described (*5*) to obtain 1−2 chelates per protein. Briefly, a protein (5−50 μM) was exchanged into a nonnucleophilic buffer (10 mM MOPS, pH 8, 100 mM NaCl, 0.1 mM EDTA, 5% glycerol) by gel filtration. FeBABE was added to the protein (final concentration 1.4 mM), immediately followed by 2IT (final concentration 0.7 mM). The reaction was incubated at 37 °C for 1 h, and excess reagents were removed by gel filtration. Glycerol was added to 50%, and the conjugates were stored at −70 °C. Conjugation yields were estimated according to ref *5*. The concentrations of the resulting conjugates were determined by a Bradford dye-binding assay (Bio-Rad).

Proximity Cleavage with 2IT-FeBABE Lysine-Labeled Proteins. The core enzyme (10 pmol) was incubated in a 1:1 molar ratio with a 2IT-FeBABE conjugate in cleavage buffer (10 mM MOPS, pH 8, 120 mM NaCl, 10 mM $MgCl_2$, 1 mM EDTA, 10% glycerol) for 15 min at room temperature (8 μL total volume). The presence of 10% glycerol, a hydroxyl radical scavenger, limits the distance a diffusible hydroxyl radical can travel. To activate the cleavage reaction, ascorbate and hydrogen peroxide (1 μL each) were added to final concentrations of 4 mM. The reaction was allowed to proceed for 1 min, followed by the addition of 2.5 μL of

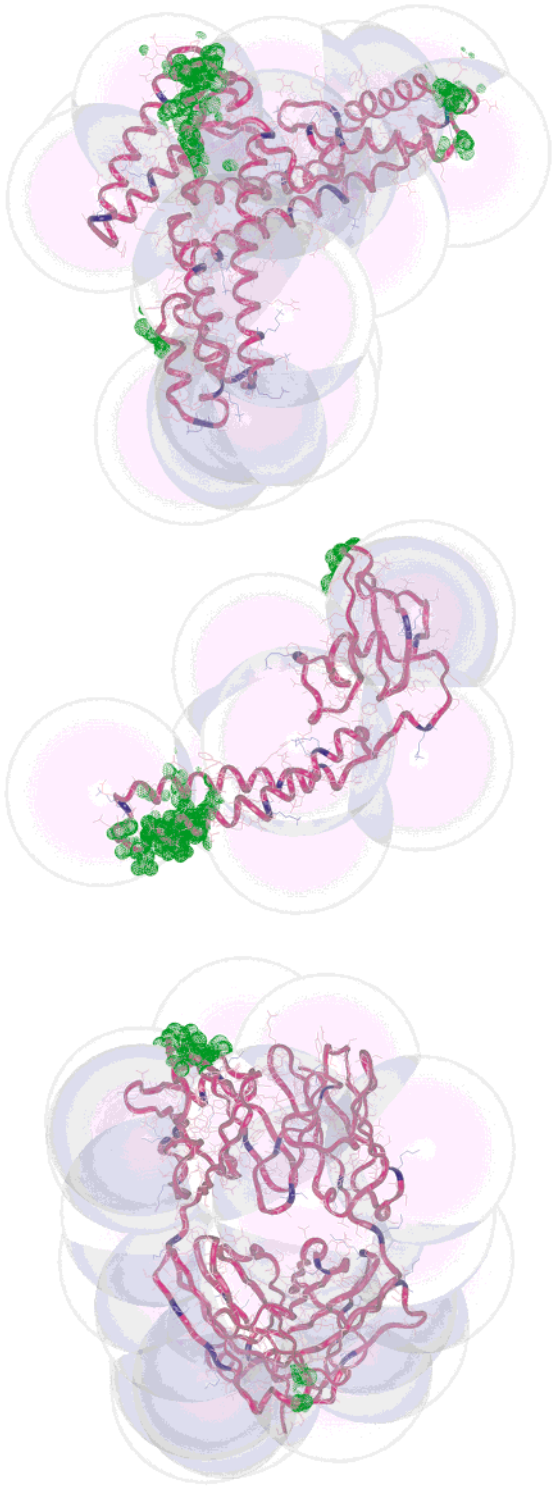

FIGURE 2: Three examples of lysine-labeled proteins: the σ^{70} fragment (*8*) at the top, GreA (*9*) in the center, and an antibody Fab fragment at the bottom (*10*). Blue segments of the ribbon are lysine residues. To depict the surface coverage of each 2IT-FeBABE lysine-labeled protein preparation (*5*), spheres with radius 18 Å (the reach of the cutting reagent) are superimposed on all the lysine ϵ-amino groups. This is a conservative estimate that does not take into account the flexibility of the lysine side chain. The minor regions of the surface *not* covered by the spheres are shown in green; molecules are oriented to emphasize these.

5× sample application buffer [62.5 mM Tris·HCl, pH 6.8, 2% (w/v) SDS, 10% (v/v) glycerol, 5% (v/v) 2-mercaptoethanol, 0.02% (w/v) bromophenol blue, 25 mM EDTA].

The samples were frozen in liquid nitrogen and stored at −70 °C.

Analysis of Cut Sites. The cleavage fragments were separated by SDS−PAGE, electrophoretically blotted onto poly(vinylidene difluoride) (PVDF) membrane, and visualized with affinity-purified antibodies against the termini of the core enzyme subunits, α, β, and β', as described by Greiner et al. (*20*). The sizes of the fragments, in terms of residue numbers, were determined by comparison to markers created by 2-nitro-5-thiocyanobenzoic acid (NTCB) and cyanogen bromide (CNBr) digests of the core enzyme (*21*) and measured using a Fluor-S MultiImager with Quantity One software (Bio-Rad). The experiments were reproduced and analyzed five times. To estimate the precision of these assignments, the number of residues in each fragment was measured from its mobility on each of the five blots. The standard deviation of these measurements was calculated for each fragment, and then these values were averaged to yield a measure of the overall precision for either β or β'.

RESULTS AND DISCUSSION

Regions within the different regulatory proteins that are involved in core RNAP binding have been previously identified, but there is less information about the complementary sites on the much larger $\alpha_2\beta\beta'$ core. The core binding site of σ^{70} has been studied previously with techniques such as chemical cross-linking (*22*), FeEDTA footprinting (*20*, *23*), and chemical cleavage with reagents tethered at specific residues of σ^{70} mutants (*21*). Cross-linking has revealed subunit−subunit proximity, but the locations of cross-links are rarely identified. FeEDTA footprinting has been used to study the binding of σ^{70} to $\alpha_2\beta\beta'$, but is problematic with weaker binders such as σ^{E}, σ^{38}, or NusA. It is also important to note that FeEDTA footprinting cannot distinguish between binding sites and conformational changes. Tethered cutting reagents attached to engineered single-cysteine mutants of σ^{70} have been used to identify sites on $\alpha_2\beta\beta'$ proximal to the location of each cutter, without interference from weak binding or conformational changes (*21*). Targeted protein footprinting is an efficient technique for comparing regions near binding sites, allowing the present survey of several regulatory proteins. It should be noted that the actual residues at the binding interface would not be revealed by this experiment; tethered reagents are not expected to cut within the binding site. However, tethered reagents can cut sites within ≈20 Å of their tether, revealing spatial proximity.

Figure 3 shows experimental results from cleavage of the β and β' subunits with lysine-labeled proteins. No cleavage products were observed on the α subunit. BSA was also conjugated with 2IT-FeBABE to demonstrate the specificity of the cleavage. The lysine-labeled BSA, which does not bind core RNAP, did not cleave it (Figure 3, lane 3). For comparison, the unbound, nonspecific cutting reagent FeEDTA (Figure 3, lane 11) produces a large number of fragments, in contrast to the limited cutting patterns of 2IT-FeBABE lysine-labeled proteins.

Figure 4 shows the locations of the cut sites along the polypeptide sequences of subunits β and β'. Primary structure segments labeled with capital letters are conserved sequences, common to RNA polymerases in many organisms (*24*). The

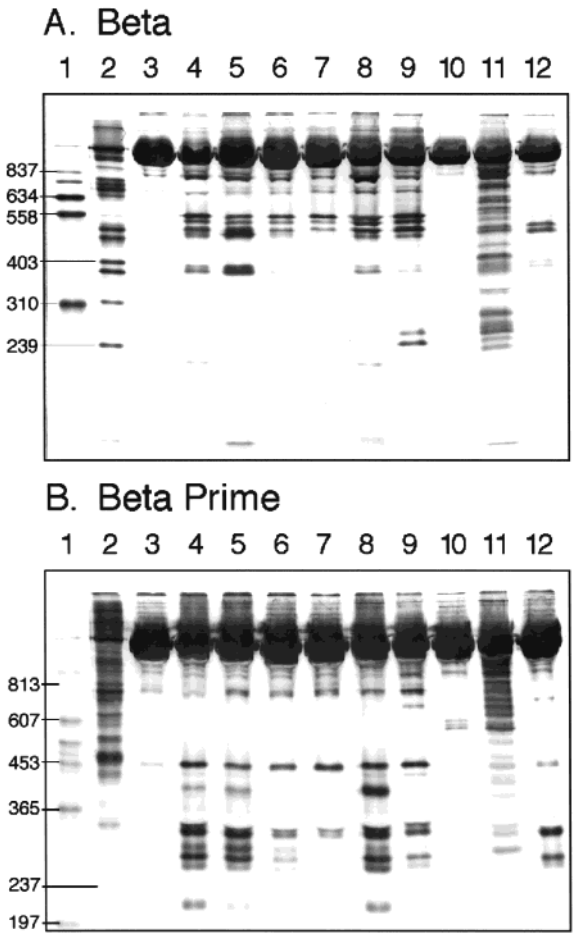

FIGURE 3: Analysis of targeted protein footprinting by SDS−PAGE, western blotting, and visualization with specific antibodies against the N-terminus of subunit β (A) or β' (B). Marker lanes 1 and 2 are the NTCB and CNBr chemical digests of $\alpha_2\beta\beta'$. Cleavage of $\alpha_2\beta\beta'$ with lysine-labeled proteins: lane 3, BSA-2IT-FeBABE; lane 4, σ^{70}-2IT-FeBABE; lane 5, σ^{54}-2IT-FeBABE; lane 6, σ^{38}-2IT-FeBABE; lane 7, σ^{E}-2IT-FeBABE; lane 8, NusA-2IT-FeBABE; lane 9, GreA-2IT-FeBABE; lane 10, omega-2IT-FeBABE. Lane 11 is untethered FeEDTA cleavage of the core enzyme (*20*). Lane 12 is cleavage with FeBABE conjugated to the single cysteine residue 396 of a σ^{70} mutant, 396C-σ^{70}-FeBABE (*21*). The residue numbers for some of the NTCB and CNBr marker fragments are indicated on the left.

smallest sigma protein, σ^{E} (∼20 kDa), responds to extracytoplasmic stress by activating transcription of heat shock proteins. The cut sites produced by lysine-labeled σ^{E} are located in or near β regions C, D, and E and in β' regions C and D. σ^{38}, a key factor during the transition from exponential growth to the stationary phase, shows cut sites in the same regions and also in region B and near G and H of β and in region B of β'. The primary sigma factor σ^{70}, which directs the majority of gene transcription during exponential growth, shows the same cuts as σ^{38} and σ^{E}, with additional cuts near β region G and β' region A.

σ^{54}, which responds to nitrogen deprivation, is structurally and functionally distinct from the σ^{70} family, which includes σ^{38} and σ^{E} (*25*). Lysine-labeled σ^{54} cuts regions of core RNAP similar to the σ^{70} family. Even though σ^{54} has almost no sequence homology to the σ^{70} family, it evidently binds within reach of the same region on core RNAP.

The NusA protein, which is involved in regulating RNA elongation, cuts in the same regions of both β and β' as the

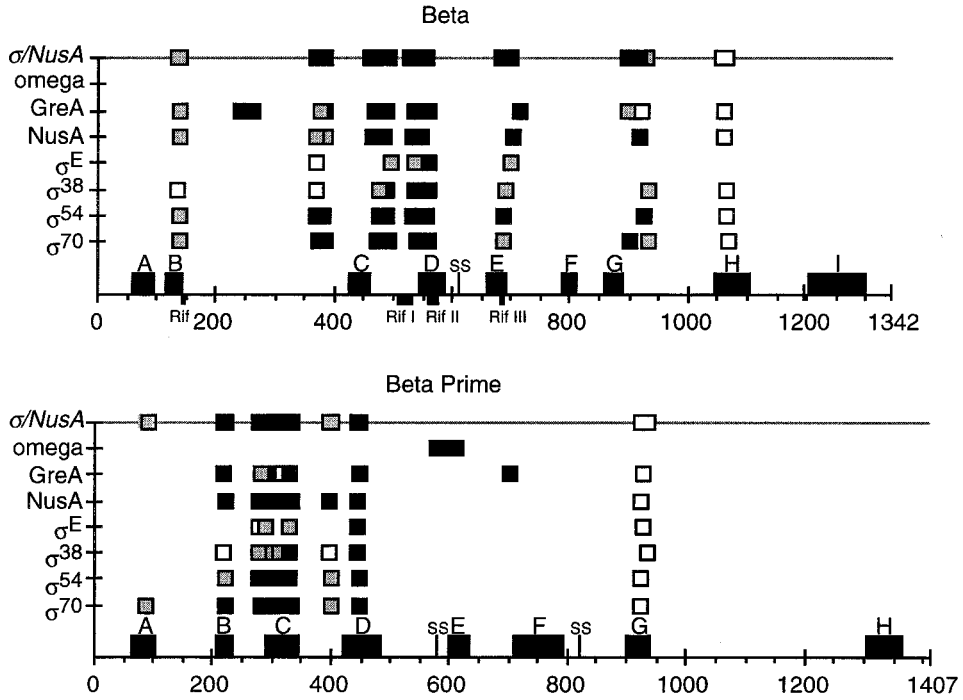

FIGURE 4: Comparison of cut sites from targeted protein footprinting with the lysine-labeled proteins indicated along the vertical axes. The accuracy of these assignments is estimated to be ±6 residues for β and ±11 for β' (*21*); the precision is ±5 residues for both. The primary sequences, conserved regions, rifampicin binding regions (Rif), and chloroplast and archaebacterial split sites (SS) of subunits β and β' are indicated along the horizontal axes. The shading of these boxes corresponds to the intensity of the cut sites. *σ/NusA* is a summary of cut sites produced by the sigma proteins and NusA.

σ factors. This is consistent with evidence that NusA and σ^{70} bind competitively to RNAP (*4*).

GreA is a transcription elongation factor that induces transcript cleavage in stalled complexes. Lysine-labeled GreA produces β cut sites similar to the other proteins but also has two unique cuts near β residues 244 and 260. Lysine-labeled GreA produces some β' cuts similar to the other proteins but makes a unique cut in β' region F. The binding of streptolydigin, which inhibits RNA elongation, is affected by mutations in this region (*26*). In eukaryotic RNA polymerase II, mutations in the corresponding region F affect the binding of α-amanitin, which inhibits nascent RNA cleavage by transcription elongation factor SII (*27*). The crystal structure of GreA shows that it consists of two distinct domains. The coiled-coil N-terminal domain juts out 45 Å in solution and plays a direct role in the transcript cleavage reaction. The lysine-labeled N-terminal domain of GreA should be within reach of the active site of RNAP and capable of cleaving similar regions of RNAP as the sigma factors. The globular C-terminal domain is required for GreA binding to RNAP. The lysine-labeled C-terminal domain of GreA is likely to have produced the unique cut site in β' region F.

The omega protein binds to and copurifies with RNAP but has no established function. Omega can be cross-linked to β' but not β (*28*). Lysine-labeled omega does not cut β but produces unique cut sites near β' region E, where a chloroplast split site is located. The large subunits of bacterial RNA polymerase remain functional after cleavage at chloroplast split sites, with one exception (*29*): a defect in promoter clearance results from cleavage of β' at the site near where lysine-labeled omega cuts.

Aside from minor differences, lysine-labeled σ^{70}, σ^{54}, σ^{38}, σ^{E}, and NusA (collectively referred to as σ/NusA) produce cut sites in the same regions of both β and β', outlining a common surface. All these proteins are involved in the control of gene expression, as is GreA, which also cuts in similar regions. We infer that these proteins bind within reach of the active site of RNA polymerase. The omega protein evidently binds elsewhere.

On β, σ/NusA cut sites are intense and clustered around the rifampicin binding regions. Rifampicin inhibits RNAP, and mutational analysis of the residues located in the rifampicin binding site indicates that they comprise part of the catalytic center (*26*). These cut sites are also just outside the conserved motif NADFDGD, which is involved in chelating the active center Mg^{2+} (*30*). These σ/NusA regions are distinct from the segments of β implicated in binding the α subunit [residues 737−904 and downstream of 1138 (*31*)].

On the β' subunit, intense σ/NusA cut sites lie primarily in region C. These cut sites are clustered between β' residues 201−477; mutant β' proteins lacking this region retain the ability to form the core enzyme but cannot bind σ (*32*). Recently, the $\sigma-\beta'$ interaction was mapped to a peptide containing β' residues 260−309 by affinity techniques (*33*). Further, residues located in β' region C are homologous to the DNA-binding cleft of DNA polymerase I (*24*). σ also cuts β' region D; β' regions A−B, E−F, and G−H are involved in binding $\alpha_2\beta$, but regions C and D are not necessary (A. Katayama, N. Fujita, and A. Ishihama, submitted for publication).

Targeted protein footprinting is a powerful tool for mapping interactions within macromolecular complexes and for comparing the binding sites of several proteins to a common target. Because it (1) reports on proximity to the protein of interest, (2) surveys practically the entire surface

of each cutting protein, and (3) provides a straightforward readout of sequence positions cleaved on the target macromolecules, this approach is a valuable adjunct to other techniques for studying macromolecular architecture. The concept can be extended by using other cutting reagents with longer or shorter ranges or with different attachment sites. Because its chemistry is versatile (*34*), this technique can be utilized to study protein−nucleic acid interactions as well. We envision that it can be applied generally to other complex systems.

ACKNOWLEDGMENT

We thank Reiko Miyake, Jeffrey Owens, Sydney Kustu, Caroline Kane, and Carol Gross for helpful discussions and gifts of material.

REFERENCES

1. Miyake, R., Murakami, K., Owens, J. T., Greiner, D. P., Ozoline, O. N., Ishihama, A., and Meares, C. F. (1998) *Biochemistry 37*, 1344−1349.
2. Ghaim, J. B., Greiner, D. P., Meares, C. F., and Gennis, R. B. (1995) *Biochemistry 34*, 11311−11315.
3. Murakami, K., Owens, J. T., Belyaeva, T. A., Meares, C. F., Busby, S. J. W., and Ishihama, A. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 11274−11278.
4. Gill, S. C., Weitzel, S. E., and von Hippel, P. H. (1991) *J. Mol. Biol. 220*, 307−324.
5. Traviglia, S. L., Datwyler, S. A., and Meares, C. F. (1999) *Biochemistry 38*, 4259−4265.
6. Lavoie, B. D., and Chaconas, G. (1993) *Genes Dev. 7*, 2510−2519.
7. Aki, T., and Adhya, S. (1997) *EMBO J. 16*, 3666−3674.
8. Malhotra, A., Severinova, E., and Darst, S. A. (1996) *Cell 87*, 127−136.
9. Stebbins, C. E., Borukhov, S., Orlova, M., Polyakov, A., Goldfarb, A., and Darst, S. A. (1995) *Nature 373*, 636−640.
10. Cohen, G. H., Sheriff, S., and Davies, D. R. (1996) *Acta Crystallogr. D 52*, 315−326.
11. Hager, D. A., Jin, D. J., and Burgess, R. R. (1990) *Biochemistry 29*, 7890−7894.
12. Fujita, N., and Ishihama, A. (1996) *Methods Enzymol. 273*, 121−130.
13. Popham, D., Keener, J., and Kustu, S. (1991) *J. Biol. Chem. 266*, 19510−19518.
14. Tanaka, K., Takayanagi, Y., Fujita, N., Ishihama, A., and Takahashi, H. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 8303.
15. Rouviere, P. E., De Las Penas, A., Mecsas, J., Lu, C. Z., Rudd, K. E., and Gross, C. A. (1995) *EMBO J. 14*, 1032−1042.
16. Olins, P. O., Erickson, B. D., and Burgess, R. R. (1983) *Gene 26*, 11−18.
17. Borukhov, S., and Goldfarb, A. (1996) *Methods Enzymol. 274*, 315−326.
18. Gentry, D. R., and Burgess, R. R. (1990) *Protein Expression Purif. 1*, 81−86.
19. Greiner, D. P., Miyake, R., Moran, J. K., Jones, A. D., Negishi, T., Ishihama, A., and Meares, C. F. (1997) *Bioconjugate Chem. 8*, 44−48.
20. Greiner, D. P., Hughes, K. A., Gunasekera, A. H., and Meares, C. F. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 71−75.
21. Owens, J. T., Miyake, R., Murakami, K., Chmura, A. J., Fujita, N., Ishihama, A. and Meares, C. F. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 6021−6026.
22. Hillel, Z., and Wu, C. W. (1977) *Biochemistry 16*, 3334−3342.
23. Heyduk, T., Heyduk, E., Severinov, K., Tang, H., and Ebright, R. H. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 10162−10166.
24. Allison, L. A., Moyle, M., Shales, M., and Ingles, C. J. (1985) *Cell 42*, 599−610.
25. Helmann, J. D. (1994) *Transcription: Mechanisms and Regulation*, pp 1−17, Raven Press, Ltd., New York.
26. Jin, D. J., and Zhou, Y. N. (1996) *Methods Enzymol. 273*, 300−319.
27. Izban, M. G., and Luse, D. S. (1992) *Genes Dev. 6*, 1342−1356.
28. Gentry, D. R., and Burgess, R. R. (1993) *Biochemistry 32*, 11224−11227.
29. Severinov, K., Mustaev, A., Kukarin, A., Muzzin, O., Bass, I., Darst, S. A., and Goldfarb, A. (1996) *J. Biol. Chem. 271*, 27969−27974.
30. Zaychikov, E., Martin, E., Denissova, L., Kozlov, M., Markovtsov, V., Kashlev, M., Heumann, H., Nikiforov, V., Goldfarb, A., and Mustaev, A. (1996) *Science 273*, 107−109.
31. Nomura, T., Fujita, N., and Ishihama, A. (1999) *Biochemistry 38*, 1346−1355.
32. Luo, J., Sharif, K. A., Jin, R., Fujita, N., Ishihama, A., and Krakow, J. S. (1996) *Genes Cells 1*, 819−827.
33. Arthur, T. M., and Burgess, R. R. (1998) *J. Biol. Chem. 273*, 31381−31387.
34. Heilek, G. M., and Noller, H. F. (1996) *Science 272*, 1659−1662.

BI9917232