Articles

Importance of Minor-Groove Contacts for Recognition of DNA by the Binding Domain of Hin Recombinase†

James P. Sluka,‡ Suzanna J. Horvath,§ Anna C. Glasgow,§ Melvin I. Simon,§ and Peter B. Dervan*,‡

Division of Chemistry and Chemical Engineering and Division of Biology, California Institute of Technology, Pasadena, California 91125

Received January 5, 1990; Revised Manuscript Received March 6, 1990

ABSTRACT: Incorporation of the DNA-cleaving moiety EDTA·Fe at discrete amino acid residues along a DNA-binding protein allows the positions of these residues relative to DNA bases, and hence the organization of the folded protein, to be mapped by high-resolution gel electrophoresis. A 52-residue protein, based on the sequence-specific DNA-binding domain of Hin recombinase (139–190), with EDTA at the NH_2 terminus cleaves DNA at Hin recombination sites. The cleavage data for EDTA–Hin(139–190) reveal that the NH_2 terminus of Hin(139–190) is bound in the minor groove of DNA near the symmetry axis of Hin-binding sites [Sluka, J. P., Horvath, S. J., Bruist, M. F., Simon, M. I., & Dervan, P. B. (1987) *Science 238*, 1129]. Six proteins, varying in length from 49 to 60 residues and corresponding to the DNA-binding domain of Hin recombinase, were synthesized by solid-phase methods: Hin(142–190), Hin(141–190), Hin(140–190), Hin(139–190), Hin(135–190), and Hin(131–190) were prepared with and without EDTA at the NH_2 termini in order to test the relative importance of the residues Gly^{139}-Arg^{140}-Pro^{141}-Arg^{142}, located near the minor groove, for sequence-specific recognition at five imperfectly conserved 12-base-pair binding sites. Footprinting and affinity cleaving reveal that deletion of Gly^{139} results in a protein with affinity and specificity similar to those of Hin(139–190) but that deletion of Gly^{139}-Arg^{140} affords a protein with altered affinities and sequence specificities for the five binding sites. It appears that Arg^{140} in the DNA-binding domain of Hin is important for recognition of the 5′-AAA-3′ sequence in the minor groove of DNA. Our results indicate modular DNA and protein interactions with two adjacent DNA sites (major and minor grooves, respectively) bound on the same face of the helix by two separate parts of the protein.

X-ray crystal structures of DNA-binding proteins and protein–DNA complexes have provided insights into the variety of discrete, organized, and tightly packed three-dimensional structures available for sequence-specific recognition of double-helical DNA. Most protein–DNA complexes characterized to date involve major-groove contacts. High-resolution crystallographic views of repressor–operator complexes utilizing a helix–turn–helix motif reveal the structural complexity of protein–DNA interactions (Anderson et al., 1985, 1987; Aggarwal et al., 1988; Otwinowski et al., 1988; Jordan & Pabo, 1988). The protein–DNA interface includes protein contacts to the sugar–phosphate backbone as well as to base pairs in the major groove. A particular side chain can contact several base pairs, and several side chains can cooperate to recognize a single base. Moreover, sequence-dependent ability of DNA to adopt the required conformation appears to be important for site-specific recognition (Otwinowski et al., 1988; Jordan & Pabo, 1988; Aggarwal et al., 1988).

Many sequence-specific DNA-binding proteins do not make use of a helix–turn–helix motif. Of these proteins, only *Eco*RI has been characterized by X-ray crystallographic analysis as a DNA complex (McClarin et al., 1986). The endonuclease binds to DNA as a dimer. Recognition occurs through the NH_2 termini of four α-helices, which are wedged into the major groove and interact primarily through hydrogen bonding to the six-base-pair recognition sequence. Other sequence-specific DNA-binding proteins that have not yet been solved by X-ray diffraction methods and are apparently different from the repressor helix–turn–helix and the endonuclease "double-barreled helix" motifs include the "zinc binding fingers" (Miller et al., 1985; Berg, 1988; Pârraga, 1988, 1990; Lea et al., 1989) and the "scissor-grip leucine zipper" (Landschulz et al., 1988; Vinson et al., 1989; Oakley & Dervan, 1990). In these cases, models of the protein–DNA structures are derived from solution studies using combinations of physical, biochemical, and genetic methods. Models for the zinc binding fingers and the scissor-grip leucine zipper dimers involve predominantly major-groove contacts.

Minor-groove protein–DNA contacts conferring sequence-specific affinity are less well documented, although a few recent examples exist. Chemical footprinting studies support a model for the *Escherichia coli* integration host factor (IHF) in which a dimer of nonidentical subunits, utilizing a pair of two-stranded β ribbon arms, interacts primarily with the minor groove of DNA (Yang & Nash, 1989). The NH_2 terminus of sea urchin spermatogenetic histone H1 and H2B, which comprises repeats of Ser-Pro-Lys(Arg)-Lys(Arg), is proposed to form a compact β-turn and form specific hydrogen bonds with the floor of the minor groove, in a manner similar to netropsin (Suzuki, 1989).

In this work, we utilize the technique of affinity cleaving

† Supported by grants from the DARPA University Research Initiative Program, the National Foundation for Cancer Research, and the National Science Foundation and by a National Research Service Award from NIGMS to J.P.S.

* To whom correspondence should be addressed.
‡ Division of Chemistry and Chemical Engineering.
§ Division of Biology.

0006-2960/90/0429-6551$02.50/0

to map the position of the NH_2-terminal end of the DNA-binding domain of Hin recombinase relative to five imperfectly conserved 12-base-pair DNA binding sites. By incrementally deleting amino acids at the NH_2-terminal end of a 52-residue protein, we specifically address the role of protein contacts in the minor groove for the sequence-specific recognition of DNA.

Mapping the Folded Structures of Proteins by Affinity Cleaving. Attachment of EDTA-Fe to a DNA-binding moiety creates a DNA-cleaving molecule that functions at physiologically relevant pH, temperature, and salt conditions (Hertzberg & Dervan, 1982, 1984). The cleavage reaction can be initiated by addition of a reducing agent such as dithiothreitol or sodium ascorbate. If the DNA-binding molecule is sequence specific, the EDTA-Fe cleaves at highly localized sites on DNA restriction fragments and plasmids (Schultz et al., 1982; Taylor et al., 1984; Dervan, 1986; Dreyer & Dervan, 1985; Sluka et al., 1987; Moser & Dervan, 1987). Because the cleaving moiety is not sequence specific, the cleavage specificity is determined only by the binding specificity of the molecule being investigated. EDTA·Fe-equipped DNA-binding molecules cleave DNA by oxidation of the deoxyribose backbone via a diffusable oxidant, presumably hydroxyl radical (Hertzberg & Dervan, 1982, 1984; Taylor et al., 1984; Dervan, 1986, Tullius, 1987). Cleavage of both DNA strands is observed and typically extends over 4–6 base pairs (Schultz et al., 1982; Taylor et al., 1984; Dervan, 1986). Due to the right-handed nature of double-helical DNA, the groove in which the EDTA·Fe is located can be identified by cleavage pattern analysis. An asymmetric cleavage pattern with maximal cleavage shifted to the 5′ or 3′ side on opposite strands corresponds to an EDTA·Fe location proximal to the major or minor groove, respectively.

Affinity cleaving has been used to study the sequence-specific recognition of double-helical DNA by naturally occurring DNA-binding antibiotics (Schultz et al., 1982; Taylor et al., 1984), designed peptide analogues for the minor groove (Youngquist & Dervan, 1985; Wade & Dervan, 1987; Griffin & Dervan, 1987), oligonucleotide-directed cleavage of single-stranded DNA (Dreyer & Dervan, 1985), and oligonucleotide-directed recognition of the major groove by triple helix formation (Moser & Dervan, 1987; Strobel et al., 1988; Povsic et al., 1989; Griffin & Dervan, 1989). From these studies, which involve less complicated DNA-binding motifs, numerous cleavage patterns caused by a diffusible oxidant in either the minor or the major groove of duplex DNA have been analyzed. This data base allows interpretation of the affinity cleaving results from conformationally more complex proteins.

Incorporation of EDTA·Fe at discrete amino acid residues within a protein allows the position and groove location of the modified residues to be mapped to nucleotide resolution. The secondary and tertiary structures of DNA-binding proteins can be analyzed by affinity cleaving by using two approaches. First, the amino acid location of the EDTA on a polypeptide chain of constant length may be varied in order to reveal key topological features of the protein–DNA complex (e.g., location of NH_2 versus COOH terminus) (Sluka et al., 1987; Mack et al., 1990). Alternatively, the length of the polypeptide may be incrementally changed while the EDTA is kept at the same terminus to examine the influence of substructures on the binding affinity of the protein, the base and groove location of the modified protein position, and the conformational flexibility of a specific peptide segment. Previous work (Sluka, 1987) and the accompanying paper (Mack et al., 1990) use the former strategy to trace the path of the folded peptide chain along its recognition site. Here, the latter approach is used to study the interactions between the residues at the NH_2 terminus of the DNA-binding region of Hin and its DNA recognition site.

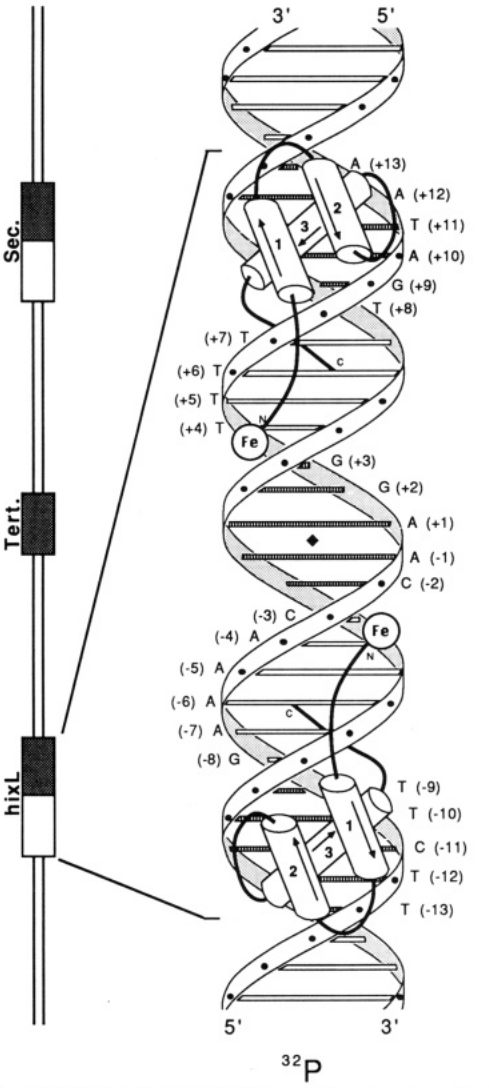


FIGURE 1: Model for Hin(139–190) and, by extension, the DNA-recognition domain of Hin binding to *hix*L (Sluka et al., 1987). The sequence of *hix*L is shown along the 5′-strand. The pseudo-C_2 axis is indicated by a solid diamond. Putative α helices are shown as cylinders with an arrow pointing from the NH_2 to the COOH terminus. The model is based on the 434 repressor fragment–14-bp operator cocrystal (Aggarwal et al., 1988) with the following changes: (i) the α_4 helix of 434 repressor is replaced with the 10-residue carboxy-terminal tail of Hin, and (ii) the nine-residue amino tail of Hin-(139–190), which includes EDTA, is added as an extension to the 434 repressor α_1 helix amino terminus. The orientation of the recognition helix has been determined by Mack et al. (1990). A schematic representation of the Hin-binding sites on the restriction fragment used in the high-resolution gel electrophoresis studies is shown on the left.

Hin Recombinase. Hin recombinase is an 190 amino acid enzyme that inverts a segment of DNA which controls the expression of the flagellin genes of *Salmonella typhimurium* (Zieg et al., 1977). Recombination occurs between two crossover sites, designated *hix*L and *hix*R, on supercoiled DNA (Johnson et al., 1984). Each *hix* site is 26 bp long and has nearly 2-fold symmetry. Hin binds to a *hix* site as a dimer and protects bases −13 through +13, inclusive, from chemical cleavage reagents [consensus half-site: 5′-TTNTCNNA-AACCA-3′ (Johnson et al., 1984; Glasgow et al., 1989a)]. A synthetic 52 amino acid protein identical with the COOH terminus of Hin (residues 139–190) has been shown by DNase

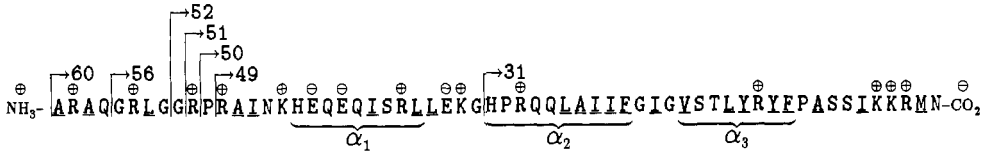


FIGURE 2: Amino acid sequences of the synthetic proteins corresponding to the DNA-binding domain of Hin recombinase. Each protein was synthesized from the COOH terminus to the indicated position. Hydrophobic residues are underlined, and charged residues are marked with the appropriate sign. With the exception of Hin 31mer, each protein was prepared with and without EDTA-GABA at the amino terminus.

I and dimethyl sulfate protection experiments to contain the sequence-specific DNA-binding activity of Hin (Bruist et al., 1987) (Figure 1). This protein fragment, Hin(139–190), protects the same DNA sequence as does the complete Hin protein except for the central three bases from −2 to +1.

The 52-residue DNA-binding region of Hin contains extensive sequence similarity to the helix–turn–helix DNA-binding proteins (Pabo & Sauer, 1984) and is thought to recognize DNA through this motif in the major groove. However, dimethyl sulfate footprinting experiments have shown that Hin recombinase and Hin(139–190) protect residues A(−6), A(−5), A(−4), A(4), A(5), and A(6) of *hix*L from methylation. These results have been taken as evidence that protein contacts also occur in the minor groove (Bruist et al., 1987, Glasgow et al., 1989a). From affinity cleaving studies, a synthetic 52-residue protein based on the DNA-binding Hin(139–190) equipped with EDTA·Fe at the amino terminus cleaves DNA near the symmetry axis of Hin recombination sites (Sluka et al., 1987). The cleavage pattern is shifted to the 3′ side, suggesting that the NH_2 terminus of Hin(139–190) is bound near or in the minor groove of DNA (Sluka et al., 1987). As the bulk of the DNA-binding domain is likely located in the adjacent major groove as a helix–turn–helix motif, residues close to the NH_2 terminus must extend across the DNA phosphodiester backbone to the minor groove. The four-residue segment at the NH_2 terminus of Hin(139–190), Gly^{139}-Arg^{140}-Pro^{141}-Arg^{142}, contains two basic side chains and could be involved in such interactions. Comparisons with the COOH-terminal domains from the closely related recombinases Gin (138–193), Pin (138–184), and Cin (138–186), as well as their DNA-binding sites, reveals that similar Arg-Gly protein sequences and DNA sequences are conserved (Glasgow et al., 1989b; Mack et al., 1990). It seems plausible that the role of this section of the protein could be to promote groove- and sequence-specific DNA binding.

Modeling the Folded Structure of the Hin DNA-Binding Domain. A binding model for the DNA-recognition domain of Hin might therefore involve at least two sets of interactions: (i) a helix–turn–helix structure binding in the major groove and (ii) a region that connects the helix–turn–helix domain with the other 140 residues of Hin by following the adjacent minor groove sequence 5′-AAA-3′ toward the center of the *hix* site. A model that incorporates these features is shown in Figure 1. The three α helices proposed in this model are based on the α_1-α_2-α_3 helices of λ cro (Anderson et al., 1981) and the 434 repressor (Anderson et al., 1987, Aggarwal et al., 1988). The orientation of the recognition helix has been determined by recent affinity cleaving studies with EDTA attached near the COOH terminus of Hin(139–186) (Mack et al., 1990). The model also shows the NH_2 terminus of the 52mer extending along the minor groove toward the center of the dimer-binding site. As demonstrated by footprinting studies, the positions contacted by Hin(139–190) in the major grooves are separated by two turns of the DNA helix (Figure 1) (Sluka et al., 1987).

Importance of the NH_2 Terminus of Hin(139–190) for DNA Recognition. The 31-residue Hin(160–190), which contains the putative helix–turn–helix motif, is incapable of binding sequence specifically to DNA (Bruist et al., 1987). Thus the helix–turn–helix motif is necessary but not sufficient for binding in these systems, highlighting the importance of other structures in the DNA-binding domain of the protein. Sequence-specific recognition by Hin could require the third α helix, such as is found in the 434 repressor and λ cro proteins, as well as structures at the NH_2 terminus of the 52mer that bind DNA in the minor groove. In this study, we assess the importance of contacts made by the NH_2 terminus of the Hin 52mer in the minor groove of DNA. We are particularly interested in the last four residues, Gly^{139}-Arg^{140}-Pro^{141}-Arg^{142}, and their role with regard to sequence-specific contacts of Hin with DNA. We synthesized by solid-phase methods six different proteins based on the DNA-binding domain of Hin (Figure 2). These proteins differ in length from 49 to 60 residues (Table I). Each protein was synthesized with and without EDTA attached to the NH_2-terminal residue for comparative affinity cleaving and footprinting studies. Three of these proteins, Hin(140–190), Hin(141–190), and Hin(142–190), were incrementally shortened at the NH_2 terminus of Hin(139–190) 52mer to afford 51-, 50-, and 49mers, respectively. Two of these proteins, Hin(135–190) and Hin(131–190), were lengthened at the NH_2 terminus of Hin(139–190) by four and eight residues to form 56- and 60mers (Figure 2, Table I).

Table I

Hin sequence	synthetic protein size
Hin(142–190)	EDTA-Hin 49mer
Hin(141–190)	EDTA-Hin 50mer
Hin(140–190)	EDTA-Hin 51mer
Hin(139–190)	EDTA-Hin 52mer
Hin(135–190)	EDTA-Hin 56mer
Hin(131–190)	EDTA-Hin 60mer

Experimental Procedures

Materials. Manual peptide syntheses were carried out in 20-mL vessels fitted with coarse glass frits as described by Kent (1988). Automated syntheses were performed on an ABI 430A synthesizer (Kent et al., 1984, 1985), modified by the removal of in-line filters to the top and bottom of the reaction vessel, with a 20-mL Teflon/KelF reaction vessel. The synthetic protocols used were developed at the California Institute of Technology (Kent & Clark-Lewis, 1985; Clark-Lewis et al., 1986; Kent et al., 1988; Kent, 1988). Protected amino acid derivatives were purchased from Peninsula Laboratories. tBoc-L-His (DNP-protected) was obtained from Fluka. Asn-(phenylacetamido)methyl (PAM) resin, dimethylformamide (DMF), diisopropylethylamine (DIEA), dicyclohexylcarbodiimide (DCC) in dichloromethane, *N*-hydroxybenzotriazole (HOBt) in DMF, and trifluoroacetic acid (TFA) were obtained from Applied Biosystems. Dichloromethane (DCM) and methanol (HPLC grade) were purchased from Mallinckrodt, *p*-cresol and *p*-thiocresol were from Aldrich, and diethyl ether (low peroxide content) was from Baker.

Doubly distilled water was used for all aqueous reactions and dilutions. Plasmid pMFB36 DNA (Bruist et al., 1987)

was isolated from *E. coli* and purified by CsCl centrifugation. Calf thymus (CT) DNA was purchased from Sigma and was sonicated, deproteinized, and dialyzed. Enzymes were purchased from Boehringer Mannheim or New England Biolabs.

Synthesis. N^{α}-tBoc-L-amino acids were used with the following side-chain protecting groups: Arg (Tos), Asp (OBzl), Glu (OBzl), His (DNP), Lys (Cl-Z), Ser (Bzl), Thr (Bzl), and Tyr (Br-Z). Manual assembly of the protected peptide on the solid support was carried out at 25 °C via a three-step reaction cycle. First the tBoc protecting group was removed from the α-amino group of the resin-bound amino acid by using TFA (65% TFA in DCM for 1 and 15 min). The deprotected peptide resin was then neutralized with 10% DIEA in DCM (2 × 1 min). Amino acids (except Asn, Gln, and Arg) were coupled to the free α-amino group as the symmetric anhydrides. For the first coupling, the symmetric anhydride was formed in situ in DCM. DMF was added after a sufficient activation time (10 min), and the reaction was allowed to proceed for a total of 30 min. Coupling yields were determined by quantitative ninhydrin monitoring (Sarin et al., 1981) with acceptable values being ≥99.7% in the beginning of the synthesis and falling off gradually to 99% near the end. If a second coupling was necessary, the resin was neutralized with 10% DIEA in DMF. The symmetric anhydride was formed outside the reaction vessel in an absolute minimum of DCM. The solution was filtered into the vessel to remove the dicyclohexylurea, topped off with DMF, and allowed to react for 1–2 h. If the yield was not acceptable at this point, a third coupling was performed by the same protocol as the second coupling.

Asn, Gln, and Arg were coupled as the HOBt esters in DMF. The HOBt ester was formed in situ by combining the amino acid, DCC, and HOBt in DMF in the reaction vessel. Reaction times were longer due to the slower nature of the coupling reaction, typically 1–2 h. When second couplings were required for Asn and Gln, the same HOBt ester procedure was followed after neutralization with 10% DIEA in DMF. Second couplings for Arg were performed by the preformed symmetric anhydride procedure above.

In order to covalently link EDTA at a discrete amino acid location in the protein Hin(139–190), a protected EDTA derivative was prepared with protecting groups compatible with Merrifield solid-phase peptide synthesis (Sluka et al., 1987, 1990). By analogy with the carboxyl side-chain protection of glutamic and aspartic acids, three of the four carboxyl arms of EDTA were protected as benzyl esters. The fourth carboxylate of the EDTA was coupled to γ-aminobutyric acid (GABA) as a linker to minimize interference between the EDTA chelate and the secondary or tertiary structure of Hin(139–190). The resulting derivatized and protected EDTA, tribenzyl-EDTA–GABA (BEG), was suitable for coupling to the amino terminus of the protected, resin-bound protein (Figure 3). BEG was activated as an HOBt ester outside the reaction vessel by the addition of solid DCC and HOBt in a minimum of DMF for 30 min. The active ester was then added to the deprotected resin-bound protein and allowed to react for 2 h. Coupling was monitored by ninhydrin analysis, and the reaction was allowed to proceed for as long as necessary to achieve complete coupling (>97%). Though BEG is formally an amino acid dipeptide, it is suitable only for capping an amino terminus, since it does not have a primary or secondary amino group for further chain extension. The synthesis of BEG will be reported elsewhere (Sluka et al., 1990).

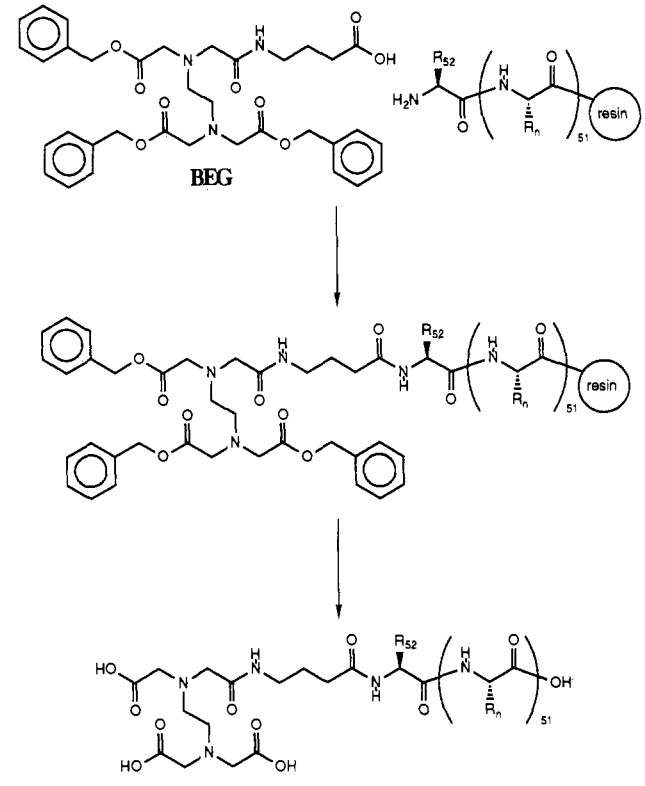


FIGURE 3: Scheme for the attachment and deprotection of tribenzyl ethylenediaminetetraacetate–GABA (BEG) to the NH_2 terminus of a protein by using solid-phase methods (Sluka et al., 1987). In the first step BEG is attached to the terminal α-amino group of a fully protected resin-bound protein in a DCC/HOBt-mediated coupling reaction. In the second step the protein is removed from the resin and the protecting groups on the amino acid side chains and the EDTA are removed by HF, affording the EDTA–protein.

Deprotection and Purification. The histidine protecting group, dinitrophenol (DNP), was removed by thiolysis at 25 °C using 20% 2-mercaptoethanol and 10% DIEA in DMF; this treatment was repeated twice for 30 min each time. After removal of the N^{α}-tBoc group with TFA and drying of the resin, all other side-chain protecting groups as well as the peptide–resin bond were cleaved by using anhydrous HF, in the presence of *p*-cresol and *p*-thiocresol as scavengers, for 60 min at 0 °C (Stewart & Young, 1981). The HF was removed under vacuum. The crude protein was precipitated with diethyl ether, collected on a fritted funnel, dissolved in water, and washed through, leaving the resin on the frit. A small sample was then removed, filtered, and subjected to analytical HPLC (Vydac 25 cm × 4.6 mm C_4 column, 0–60% acetonitrile/0.1% TFA over 60 min). The remaining solution was frozen and lyophilized. Residual DNP groups were removed from the crude protein by treatment in 4 M guanidine hydrochloride, 50 mM Tris, pH 8.5, and 20% 2-mercaptoethanol for 1 h at 50 °C (Kent, 1988). This solution was injected directly onto a semipreparative C_4 HPLC column (25 × 1 cm) and run in H_2O/0.1% TFA until the guanidine and 2-mercaptoethanol had eluted. A gradient of 0–60% acetonitrile/0.1% TFA was run over 240 min, and fractions were collected. Fractions were checked by analytical HPLC, and those containing the desired peak were pooled and lyophilized. The protein was packaged into (5×10^{-9})-mol samples on the basis of OD_{275} (ϵ = 2800 for 2 Tyr). NH_2-terminal sequence determination of the synthetic 51mer, Hin(140–190), confirmed the sequence order to be correct.

Footprinting and Affinity Cleavage Assays. The plasmid

pMFB36 (Bruist et al., 1987) was linearized by cleavage with restriction endonuclease *Xba*I. Labeling at the 3′ end was accomplished with [α-^{32}P]dATP and the Klenow fragment of DNA polymerase I. The 5′ end was labeled with ^{32}P by treatment with calf alkaline phosphatase (CAP) followed by phosphorylation with [γ-^{32}P]ATP and T4 polynucleotide kinase. Cleavage with restriction endonuclease *Eco*RI yielded 3′- and 5′-end-labeled fragments of 557 bp, which were isolated by polyacrylamide gel electrophoresis.

MPE footprinting and affinity cleavage of DNA were investigated by analysis of the cleavage patterns produced on a ^{32}P-end-labeled (both 5′ and 3′) DNA restriction fragment containing *hix*L and *hix*R separated by 400 base pairs and a closely homologous 27-bp sequence, termed the secondary Hin-binding site, which is located near *hix*L. For footprinting, the proteins without EDTA were allowed to equilibrate at micromolar concentrations (22 °C, 20 mM NaCl and 20 mM Tris, pH 8.0) with the DNA for 10 min, followed by footprinting with MPE·Fe(II) (Dervan, 1986). The cleavage products were separated by high-resolution polyacrylamide gel electrophoresis and visualized by autoradiography.

For affinity cleaving, Fe·EDTA–Hin(139–190) was prepared by mixing 1 equiv of ferric ammonium sulfate with EDTA–Hin(139–190) followed by dilution to the appropriate micromolar concentrations. The EDTA-containing proteins were allowed to equilibrate in a solution of 5′- or 3′-end-labeled DNA (22 °C, 20 mM NaCl and 20 mM Tris, pH 8.0) containing tRNA as carrier for 10 min, followed by initiation of the cleavage reaction by addition of dithiothreitol (DTT). The reactions were allowed to proceed for 40–60 min at 22 °C and terminated by ethanol precipitation. The ^{32}P-labeled DNA products were analyzed by denaturing polyacrylamide gel electrophoresis followed by autoradiography.

Gel Electrophoresis DNA Binding Assays. The gel electrophoresis DNA binding assays with 49-, 50-, 51-, 52-, 56-, and 60mer proteins were performed as described previously (Glasgow et al., 1989a). The DNA fragments used in these DNA binding studies were the 269-bp and 253-bp *Nar*I–*Hae*II restriction fragments from pMS551 (*hix*L) and pMS 570 (consensus *hix* half-site), respectively (Glasgow et al., 1989a). Reaction conditions for the binding assays were 80 mM NaCl, 20 mM Tris·HCl, pH 7.6, 1.2 mM EDTA, 5 mM $MgCl_2$, 1 mM DTT, 0.5 mg/mL bovine serum albumin, and 5 μg/mL sonicated calf thymus DNA. The indicated concentrations of 51-, 52-, 56-, and 60mers (see Results) were added to binding reactions containing $\sim 2 \times 10^{-1}$ cpm of ^{32}P-end-labeled fragment; protein concentrations up to 3.5 μM were used for the 49- and 50mers. The mole fraction of DNA bound by protein and the apparent binding constant at equilibrium binding were determined as described by Glasgow et al. (1989a,b).

Results

Syntheses. The 49-, 50-, 51-, and 52mers were synthesized by manual solid-phase techniques with optimized protocols on (phenylacetamido)methyl (PAM) resin using tBoc-protected amino acids from a single 1-mmol synthesis. The chain was extended to 52 residues with a portion of resin being removed at lengths of 49, 50, and 51 amino acids. The average yield per cycle, as determined by quantitative ninhydrin analysis of resin samples taken after each cycle (Sarin et al., 1981), over 50 couplings was 99.4%. The 56- and 60mers were synthesized by automated solid-phase techniques using modified software (Clark-Lewis et al., 1986) from a single 0.5-mmol synthesis. All residues were double-coupled and a resin sample was taken after the second coupling. The average yield per cycle, by ninhydrin analysis, over 59 couplings was 99.4%. Ninhydrin analysis indicated >99% yields for the coupling of BEG to the amino termini of the proteins in 3 h. The synthetic proteins were purified by preparative reverse-phase HPLC.

DNA Binding and Cleaving. The DNA-binding and -cleaving properties of the synthetic proteins were investigated by footprinting and affinity cleaving assays on a ^{32}P-end-labeled DNA restriction fragment containing *hix*L and a closely homologous 27-bp site termed the secondary site. MPE footprinting results are shown in Figure 4A, and histograms of the protection patterns are displayed in Figure 4B. The intensity of the patterns indicate that the 51-, 56-, and 60mers have affinities and specificities similar to those of the 52mer. The protection pattern observed for the 50- and 49mers indicates no binding at *hix*L. However, the protection patterns observed for the 50- and 49mers give comparable footprints at the secondary site at concentrations 10 times higher than the other proteins. These proteins also show binding to a new site termed the tertiary site. This site is also bound by the 52mer at the elevated concentrations used for the 49- and 50mers (data not shown).

The cleavage patterns obtained when the EDTA–proteins of varying length from the DNA-binding domain of Hin were allowed to react with the labeled restriction fragments are shown in Figure 5A. The cleavage intensities at each site are presented in histogram form in Figure 5B. The 51mer gives a cleavage pattern identical at all sites with that of the 52mer at the same concentration. The 56- and 60mers give very broad cleavage patterns at *hix*L and a pattern similar to the 52mer at the secondary site. As in the footprinting experiments, the 49- and 50mers give comparable cleavage intensities only at concentrations 10 times higher than required by the 52mer. There is no cleavage observed at *hix*L at this concentration, but considerable cleavage is seen at the secondary and tertiary sites. In all cases the cleavage patterns observed are asymmetrically shifted toward the 3′ end on opposite strands of the DNA.

Binding Affinities. Relative binding affinities of the different synthetic proteins were estimated by both gel retardation assays (Fried & Crothers, 1981; Garner & Revzin, 1981) and the affinity cleaving data. In the gel electrophoresis DNA binding assays, quantitation of the amount of complexed relative to uncomplexed DNA at equilibrium binding for a range of protein concentrations allowed the determination of apparent binding constants of the DNA-binding domains of Hin (Figure 6). Binding affinities for the Hin 52-, 56-, and 60mers are approximately 7.0×10^6, 1.0×10^6, and 2.0×10^6 M^{-1} (80 mM NaCl), respectively, as compared to the apparent binding constant for Hin recombinase, 1×10^9 M^{-1}, measured under similar conditions (Glasgow et al., 1989a,b). It was not possible to determine an association constant for the 51mer due to the instability of 51mer–*hix*L complexes to electrophoresis under conditions that gave stable DNA complexes for the 52-, 56-, and 60mers (Figure 6). The Hin 49- and 50mers showed no binding to the *hix*L sequence by this method.

Relative cleavage efficiencies for Fe·EDTA–Hin 49-, 50-, and 51mers relative to Hin 52mer were compared (Table II). If the efficiency of EDTA·Fe cleavage is the same for all proteins and is dependent only on the percentage of protein bound, the relative binding affinities of Hin 49–52mers can be estimated by comparison of the extent of DNA cleavage at different concentrations. The cleavage efficiencies of the different proteins at a given site relative to Fe·EDTA–Hin-(139–190) at that site were compared, as were the cleavage

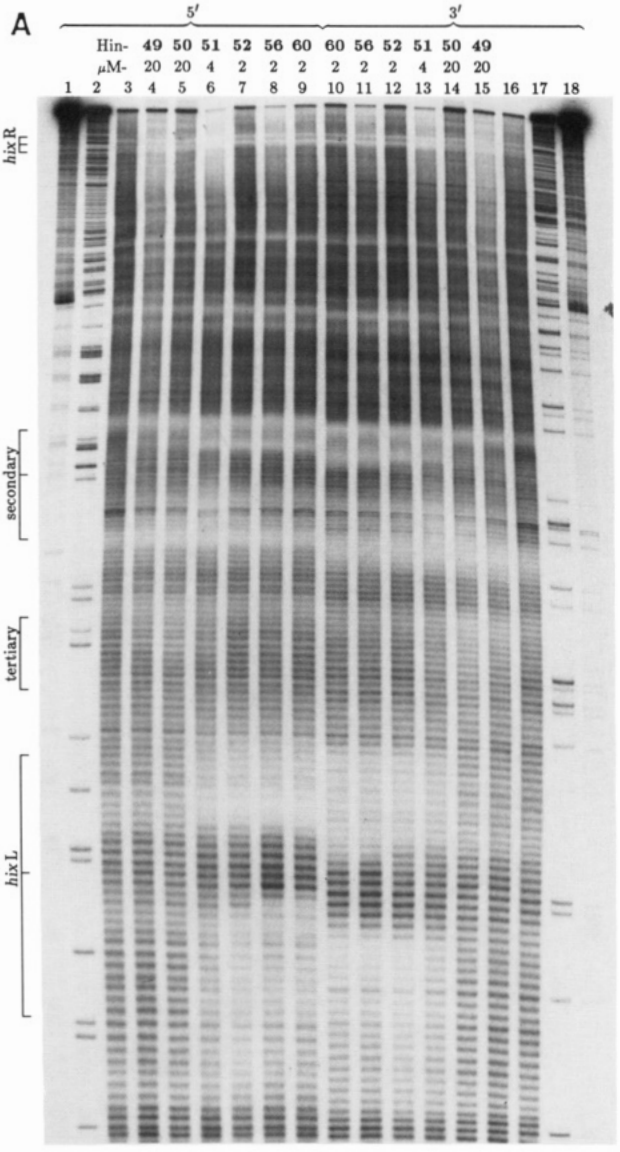
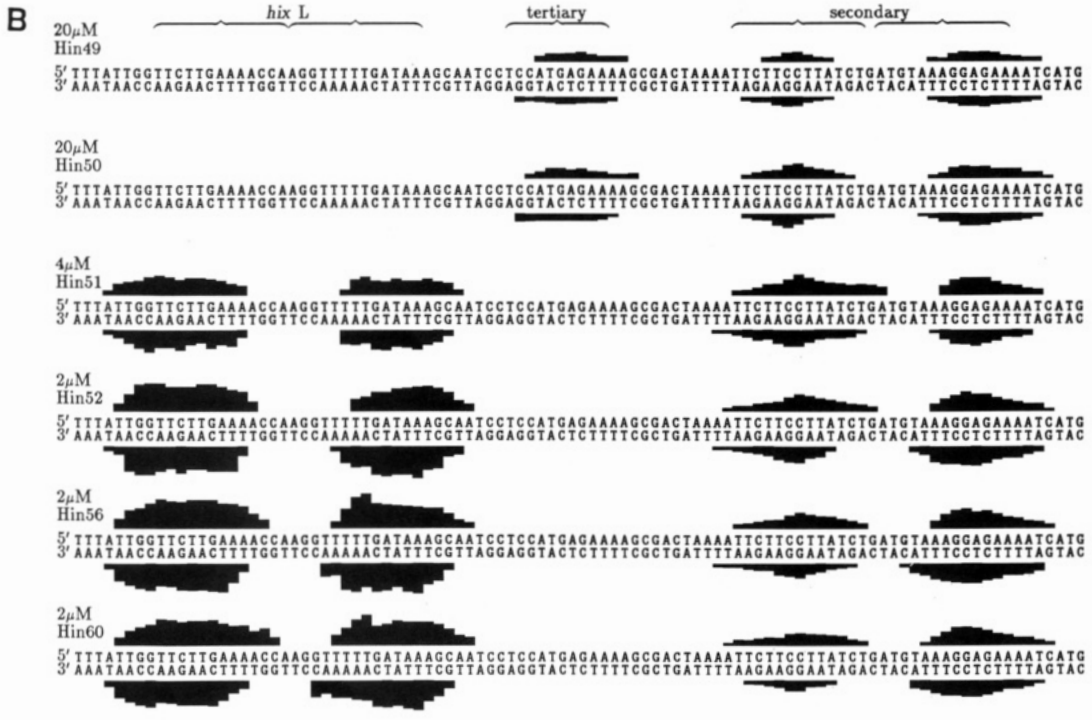


FIGURE 4: (A) Autoradiogram of high-resolution denaturing polyacrylamide gel showing MPE·Fe footprinting of Hin(142–190), Hin(141–190), Hin(140–190), Hin(139–190), Hin(135–190), and Hin(131–190) on a ^{32}P-end-labeled fragment (*Xba*I–*Eco*RI) from pMFB36 (Bruist et al., 1987). Reaction conditions were 20 mM NaCl, 29 mM Tris·HCl, pH 8.0, 400 μM tRNA, 5 mM DTT, and ≈20000 cpm (≈1 ng) of end-labeled DNA in a total volume of 15 μL. Reactions were run (without MPE equilibration) for 7.5 min at 22 °C and terminated by ethanol precipitation. Cleavage products were analyzed on an 8%, 1:20 cross-linked, 50% urea denaturing polyacrylamide gel. Lanes 1–9 contain 5′-end-labeled and lanes 10–18 contain 3′-end-labeled DNA. Lanes 1 and 18 are DNA controls, and lanes 2 and 17 are Maxam–Gilbert chemical sequencing lanes. Lanes 3–16 are footprinting reactions with 15 μM MPE·Fe: lanes 4 and 15 contain 20 μM Hin 49mer, lanes 5 and 14 contain 20 μM Hin 50mer, lanes 6 and 13 contain 4 μM Hin 51mer, lanes 7 and 12 contain 2 μM Hin 52mer, lanes 8 and 11 contain 2 μM Hin 56mer, and lanes 9 and 10 contain 2 μM Hin 60mer. (Left) Brackets mark *hix*L, tertiary, secondary, and *hix*R binding sites. (B) Histograms of the footprinting data from (A). The sequence from left to right is from the bottom of the gel to just above the secondary site. (Top) Brackets mark *hix*L, tertiary, and secondary sites. Bars are histograms to indicate extent of protection from MPE·Fe cleavage.



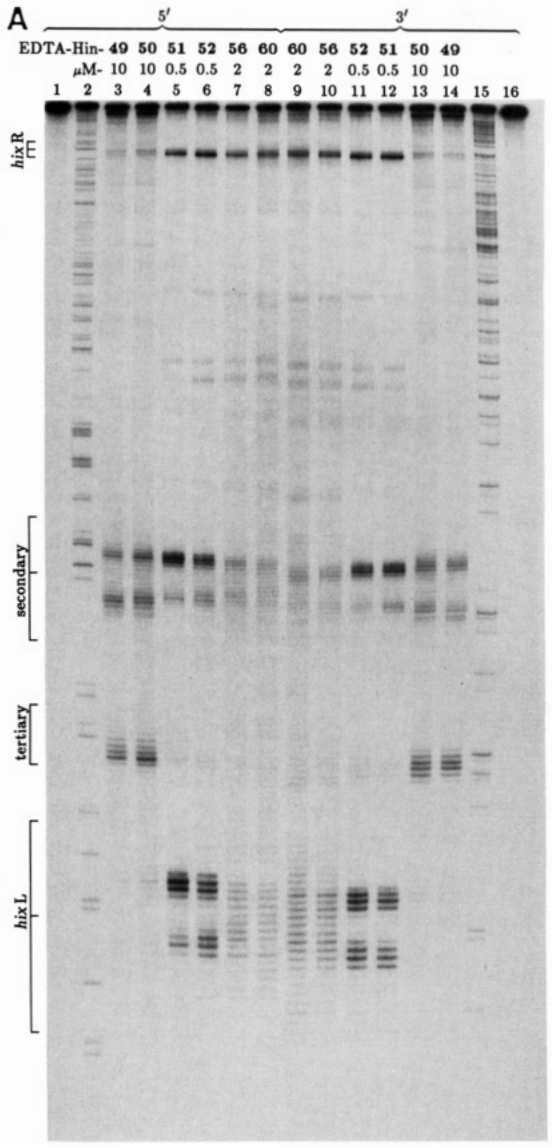


FIGURE 5: (A) Autoradiogram of a high-resolution denaturing polyacrylamide gel showing affinity cleaving reactions on a ^{32}P-end-labeled fragment (*Xba*I–*Eco*RI) from pMFB36. Reaction conditions were 20 mM NaCl, 20 mM Tris·HCl, pH 8.0, 400 μM tRNA, 5 mM DTT, and ≈15000 cpm (≈1 ng) of end-labeled DNA in a total volume of 15 μL. Reactions were run for 60 min at 22 °C and terminated by ethanol precipitation. Cleavage products were analyzed on an 8%, 1:20 cross-linked, denaturing polyacrylamide gel. Lanes 1–8 contain 5′-end-labeled DNA, and lanes 9–16 contain 3′-end-labeled DNA. Lanes 1 and 16 are DNA controls, and lanes 2 and 15 are Maxam–Gilbert chemical sequencing lanes. Lanes 3 and 14 contain 10 μM Fe·EDTA–Hin 49mer, lanes 4 and 13 contain 10 μM Fe·EDTA–Hin 50mer, lanes 5 and 12 contain 0.5 μM Fe·EDTA–Hin 51mer, lanes 6 and 11 contain 0.5 μM Fe·EDTA–Hin 52mer, lanes 7 and 10 contain 2 μM Fe·EDTA–Hin 56mer, and lanes 8 and 9 contain 2 μM Fe·EDTA–Hin 60mer. (Left) Brackets mark *hix*L, tertiary, secondary, and *hix*R binding sites. (B) Histograms of the cleavage data from (A). The sequence left to right is from the bottom of the gel to just above the secondary site in (A). (Top) Brackets mark *hix*L, tertiary, and secondary sites. Arrows indicate extent of cleavage resulting in removal of the indicated base. Dots mark the EDTA·Fe location assigned by using the model in Sluka et al. (1987) and Mack et al. (1990).

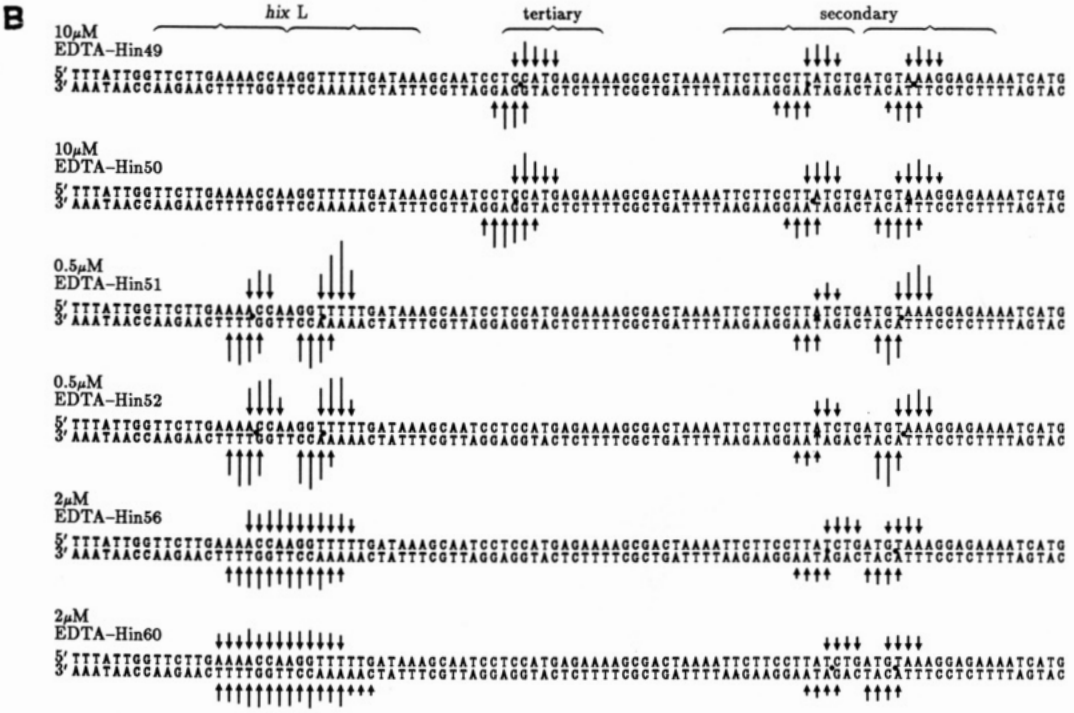


Table II: Relative Cleavage Efficiencies (RCE)[a] for EDTA–Hin 52–49mers at *hix*-like Binding Sites

half-site[b]	sequence[c,d]	52mer	51mer	50mer	49mer
	Relative to EDTA–Hin 52mer at the Same Site				
*hix*L IRL	TTCTTGAAAACC	1.0	0.7	0.003	0.001
*hix*L IRR	TTATCAAAAACC	1.0	1.1	0.007	0.007
secondary IRL	TTCTTCCTTATC	1.0	1.1	0.11	0.09
secondary IRR	TTCTCCTTTACA	1.0	1.2	0.05	0.03
tertiary	TTCTCATGGAGG	1.0	0.8	0.6	0.5
	Relative to EDTA–Hin 52mer at *hix*L IRR				
*hix*L IRL	TTCTTGAAAACC	0.9	0.6	0.003	0.001
*hix*L IRR	TTATCAAAAACC	1.0	1.1	0.007	0.007
secondary IRL	TTCTTCCTTATC	0.3	0.3	0.03	0.02
secondary IRR	TTCTCCTTTACA	0.7	0.8	0.04	0.02
tertiary	TTCTCATGGAGG	0.04	0.06	0.04	0.03

[a] Cleavage for a given compound at a particular site (A_b^a, a = compound, b = site) is the sum of the integrated peak areas over the cleavage loci on both DNA strands. RCE is the ratio of cleavage for two compounds corrected for differences in concentration (±20%). For example, RCE for EDTA–Hin 51mer at *hix*L IRR versus EDTA–Hin 52mer at the same site is RCE = $\{(A_{hixL\,IRR}^{EDTA-Hin51})/[EDTA-Hin51]\} \times \{[EDTA-Hin52]/(A_{hixL\,IRR}^{EDTA-Hin52})\}$. [b] IRL refers to inverted repeat left half-site, and IRR refers to inverted repeat right when the switching region is in the "on" configuration. [c] Sequence (5′ → 3′) toward the pseudo-C_2 axis. Boldface letters indicate base pairs likely important for minor-groove contacts (Sluka et al., 1987). [d] The consensus half-site is 5′-TTNTCNNAAACCA-3′ (Glasgow et al., 1989b).

efficiencies between the indicated protein at a particular site relative to Fe·EDTA–Hin(139–190) at the best binding site (*hix*L IRR).

Discussion

The attachment of EDTA to the NH_2 terminus of Hin(139–190) allowed us to address several issues regarding the DNA-binding domain of Hin (Figure 1). Are all 52 residues at the NH_2 terminus necessary for sequence-specific binding or do shorter proteins contain all of the required elements? What is the importance of the Gly^{139}-Arg^{140}-Pro^{141}-Arg^{142} sequence located near the minor groove of the Hin DNA site? Does the 52 amino acid sequence contain all of the important residues at the COOH terminus of Hin recombinase for sequence-specific DNA recognition?

We prepared a series of proteins that includes shorter and longer versions of the 52-residue DNA-binding domain, Hin(139–190). Since it has previously been shown that the 31-residue COOH terminus of Hin recombinase [Hin(160–190)] does not bind sequence specifically to DNA, we chose proteins ranging in size from 49 to 60 amino acids, all corresponding to the COOH terminus of Hin (Figure 2). Each protein was prepared with and without EDTA–GABA at the NH_2 terminus (Table I).

DNA Binding by Hin(135–190) 56mer and Hin(131–190) 60mer. Unlike the 52mer, both the 56- and 60mers with EDTA at the NH_2 terminus gave very broad cleavage patterns at *hix*L that cover ≈10 base pairs (Figure 5A, compare lanes 6 and 8). As determined by MPE footprinting, the 56- and 60mers have unaltered sequence specificity at the same concentration as the 52mer, indicating that these three proteins have similar binding affinities. Gel-retardation assays confirm that the association constants for the 56- and 60mers differ by less than a factor of 7 from that of the 52mer. MPE footprinting also suggests that the 52-, 56-, and 60mers all have similar affinities for the secondary and tertiary sites (Figure 4A, lanes 7–9). Thus, it appears that no additional DNA-binding features have been gained in either the 56- or 60mers. The broad cleavage patterns obtained at *hix*L with the 56- and 60mers suggest that the EDTA·Fe may not be held as close to the DNA helix as for EDTA–Hin(136–190) or that residues 131–138 may be conformationally flexible. The more localized cleavage pattern at the secondary site indicates that the secondary and *hix*L sites may have different protein–DNA interactions. These residues could comprise part of the linking region between the sequence-specific DNA-binding domain

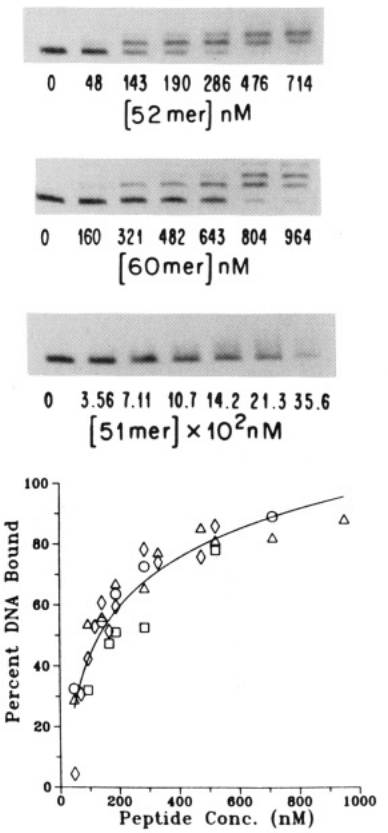


FIGURE 6: Titrations of Hin proteins' binding to the *hix* site. A plot of binding data (% DNA bound vs protein concentration) derived from titrations of the 52mer with ^{32}P-end-labeled DNA fragments from pMS551 (wild-type *hix*L) and pMS570 (*hix*L IRR) is shown. Squares and circles represent titration data from two independent experiments with wild-type *hix*L; triangles and diamonds similarly denote binding to the *hix*L IRL half-site. Protein-DNA complexes and uncomplexed DNA were separated by nondenaturing polyacrylamide gel electrophoresis as shown for 52-, 60-, and 51mer binding to the wild-type *hix*L site in the autoradiograms above the 52mer binding plot. the fastest migrating band is uncomplexed DNA; the protein–DNA complexes are retarded in their electrophoretic mobility. The values for the percentage of DNA bound in assays with the full *hix*L site included the protein–DNA complexes that had a single half-site (faster migrating complex) and both half-sites (slower migrating complex) bound by the 52mer; binding of the 52mer to the half-site alone gave a single complex of shifted mobility (data not shown). The apparent binding constant for 52mer binding to either *hix*L or the isolated *hix*L half-site, based on the protein concentration required for 50% occupancy of the *hix*L sequence, is approximately 7×10^6 M^{-1} (80 mM NaCl).

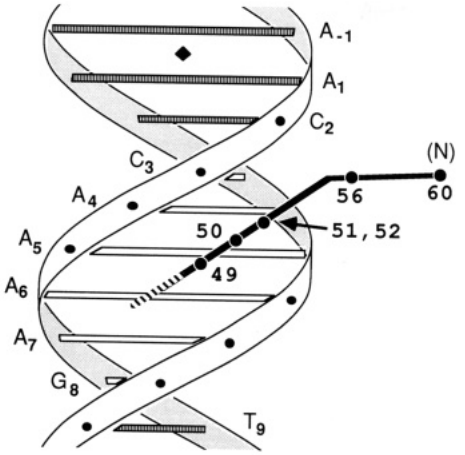


FIGURE 7: Approximate location of the residues at the NH_2 terminus of the DNA-binding domain of Hin shown by plotting the Fe·EDTA locations for EDTA–Hin 49-, 50-, 51-, 52-, 56-, and 60mers along the DNA. The DNA sequence shown along the 5′ strand is that of *hix*L IRL, and the solid diamond marks the pseudo-C_2 axis.

and the recombination and dimerization domains of the enzyme.

DNA Binding by Hin(142–190) 49mer, Hin(141–190) 50mer, and Hin(140–190) 51mer. The 51mer, which lacks a Gly^{139} residue at the NH_2 terminus, appears to have the same affinity and sequence specificity as the 52mer, as analyzed by MPE footprinting and affinity cleaving studies (Figures 4 and 5). The 50mer differs from the 52mer by deletion of Gly^{139} and Arg^{140} residues. The 49mer is lacking Gly^{139}, Arg^{140}, and Pro^{141}. The sequence-specific DNA-binding affinities of 49- and 50mers are significantly different from those of the 51- and 52mers (compare lanes 4 and 5 in Figure 5A). Cleavage at *hix*L by EDTA–Hin 49- and 50mers has been completely lost, even at 20-fold higher protein concentration. At the secondary Hin site, cleavage with Fe·EDTA–Hin 49- and 50mers is reduced by a factor of 20. The only binding site at which the 49- and 50mers have affinities comparable to those of the 51- and 52mers is the weak binding tertiary site. Footprinting of the Hin 49- and 50mers indicates that these proteins bind to the secondary and tertiary binding sites and not to the *hix*L site. The observation that the shorter proteins do bind the *secondary* and *tertiary* sites is taken as evidence that the structure of Hin remains intact and that no major structural changes occur in the putative helix–turn–helix domain of the 49- and 50mers.

Position of the NH_2 Terminus of the DNA-Binding Domain of Hin. The cleavage patterns for Fe·EDTA–Hin 49mer through Fe·EDTA–Hin 60mer are shifted to the 3′-side at all binding sites, indicating that the EDTA·Fe groups at the NH_2-terminal ends of these six proteins are located proximal to the DNA minor groove. Furthermore, the cleavage data at each half-site allow determination of the average EDTA·Fe position for each protein. A schematic representation of the average iron position versus protein length size is shown in Figure 7. The 51- and 52mers place the EDTA·Fe at the same base position. Incremental deletion of each amino acid from positions 51 to 49 shifts the EDTA·Fe location by half a base pair. The additional four amino acids of the Hin 56mer shift the EDTA·Fe an additional base pair in the opposite direction. The effect of extending four more amino acids (56–60mer) is difficult to define due to the diffuseness of the cleavage patterns.

Relative Affinities for Different Hin-Binding Sites. The first five base pairs of the five Hin-binding half-sites match the consensus half site, 5′-TT(C/A)T(T/C)-3′, while the following two base pairs are variable positions (Table II). The remaining three base pairs are proximal to the NH_2 terminus and constitute the significant differences between the sites [AAA (*hix*L), TTA (secondary), GGA (tertiary)]. The relative affinities of EDTA–Hin 52mer for the five half-sites are *hix*L IRR ~ *hix*L IRL > secondary IRR > secondary IRL ≫ tertiary, with affinities for the *hix* sites 1.5 to 3.7 times greater than for secondary sites and 30 times greater than for the tertiary site. EDTA–Hin 51- and 52mers exhibit similar binding affinities and specificities for the five half-sites. EDTA–Hin 50- and 49mers are similar to each other but show *reversed relative affinities* compared to EDTA–Hin 52mer: tertiary ~ secondary IRR ~ secondary IRL > *hix*L IRR ~ *hix*L IRL (Table II). Binding by EDTA–Hin 49- and 50mers is reduced ≥150-fold at *hix*L relative to EDTA–Hin 52mer; however, EDTA–Hin 49- and 50mer binding is reduced at the tertiary site only by a factor of 2 relative to EDTA–Hin 52mer.

The relative binding affinity of EDTA–Hin(139–190) correlates with the extent of deviation of each of the five Hin half-sites from the 5′-AAA consensus: 5′-AAA (*hix*L) > 5′-TTA (secondary) ≫ 5′-GGA (tertiary). As the protein is shortened at the amino terminus by the sequential removal of Gly^{139}, Arg^{140}, and Pro^{141} (Hin 51-, 50-, and 49mer, respectively), the effect on protein affinity for those sites with the 5′-AAA sequence is dramatically decreased. It appears then that Arg^{140}, perhaps in conjunction with the adjacent residues, is important for contacting the 5′-AAA sequence in the minor groove. The Arg^{140} could recognize a structure present in the AAA tract or make specific hydrogen-bonding contacts with the adenine residues on the floor of the minor groove.

The putative helix–turn–helix domain of Hin appears to bind specifically to the sequence TT(C/A)T(T/C) (positions −13 to −9) in the major groove at all five DNA sites studied. The tertiary site is particularly interesting, since the sequence of the adjacent minor-groove DNA binding site changes from AAA or TTA to GGA. The Hin 52-, 51-, 50-, and 49mers all bind with roughly the same (low) affinity to the tertiary site (Table II). Why, then, do the 49- and 50-residue proteins bind so poorly to the *hix*L site, while they bind almost as well as the 52mer to the secondary and tertiary sites (Table II)? This anomaly might suggest that there is some *negative* effect on Hin helix–turn–helix binding in the major groove exerted by the AAA sequence of *hix*L (perhaps due to the unusual structure adopted by adenine tracts), which is overcome by the interaction of Arg^{140} with this sequence. It is possible that Arg^{140} causes a structural alteration of the adenine tract, which influences the adjacent major-groove conformation for binding the Hin helix–turn–helix domain.

Comparison with Other Proteins. The NH_2-terminal sequence of the Hin 52mer, Gly^{139}-Arg^{140}-Pro^{141}-Arg^{142}, appears to be important for binding of the 5′-AAA tract of a *hix* half-site. Several related protein sequences from the cross-reactive recombinases Gin, Pin, and Cin (Glasgow et al., 1989b; Mack et al., 1990) might be capable of recognizing the same DNA sequence from the minor groove. The sequence Arg-Pro-(Arg/Lys) occurs in all but Gin. The substitution of Arg → Lys suggests that there may be more than one mode of recognition or that an Arg-Pro-(Arg/Lys) sequence is competent at recognition of 5′-AAA in the minor groove by a common motif.

Recombinases are not the only DNA-binding proteins that contain Gly-Arg-Pro-Arg sequences. Mammalian high mobility group-1 protein (HMG-1, also called α-protein), which binds A·T tracts of DNA, contains repeats of the sequence

Gly-Arg-Pro-Arg (Solomon et al., 1986; Lund et al., 1987). Ethylation and methylation interference studies indicate that HMG-1 binds stretches of five or more A·T base pairs via minor-groove contacts.

At least one DNA–protein complex for which high-resolution X-ray diffraction data is available contains an Arg-Pro-Arg sequence. 434 repressor contains Lys^{40}-Arg^{41}-Pro^{42}-Arg^{43} near the COOH terminus of the DNA-binding helix of its helix–turn–helix domain in a small loop between α_3 and α_4 (Aggarwal et al., 1988). This sequence interacts with the 3′ strand of the DNA binding site (ATA). Ethylation of three phosphates on this strand interferes with repressor binding. It appears that the peptide backbone makes direct hydrogen-bond contacts with the DNA–phosphate backbone at Lys^{40}, Arg^{41}, and Arg^{43}. The Arg^{43} side chain penetrates into the minor groove forming a water-mediated hydrogen bond to sugar or phosphate at the A·T base pairs at the center of the site. Arg^{43} appears to be in position for favorable electrostatic interactions with the phosphates on both strands of the minor groove, which is narrower than idealized B-form DNA. Aggarwal and co-workers suggest that the presence of a positively charged side chain in the minor groove stabilizes its compression, which brings negatively charged phosphates close together (Aggarwal et al., 1988). They point out that changing Arg^{43} to Ala decreases binding by a factor of more than 200, implying that this residue is critical for strong binding. In the 434 repressor model, then, the sequence Arg-Pro-Arg recognizes a sequence-specific DNA conformation that is characterized by a narrow minor groove. This may be relevant to the Hin case.

In conclusion, this report is one of a few examples that proteins can utilize *specific contacts within the minor groove*, as well as within the major groove, for sequence-specific recognition of duplex DNA. Our data support a model where the DNA-binding domain of Hin recombinase (residues 139–190) binds DNA specifically, utilizing contacts in adjacent minor and major grooves of each Hin-binding site. The amino acid residue Arg^{140} within the sequence Gly^{139}-Arg^{140}-Pro^{141}-Arg^{142} is important for recognition of A·T tracts in the minor groove of Hin-binding sites, while the putative helix–turn–helix domain in the adjacent major groove specifically binds the sequence 5′-TTCT(T/C)-3′. This result may be taken as evidence for modular DNA and protein interactions with two adjacent DNA sites on the same face of the helix bound cooperatively by two separate parts of the protein.

Registry No. 60mer, 127445-31-2; 56mer, 127445-32-3; 52mer, 114451-39-7; 51mer, 127445-33-4; 50mer, 127445-34-5; 49mer, 127445-35-6; 31mer, 127445-36-7; Arg, 74-79-3.

References

Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M., & Harrison, S. C. (1988) *Science 242*, 899.

Anderson, J. E., Ptashne, M., & Harrison, S. C. (1985) *Nature (London) 316*, 596.

Anderson, J. E., Ptashne, M., & Harrison, S. C. (1987) *Nature 326*, 846.

Berg, J. M. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 99.

Bruist, M. F., Horvath, S. J., Hood, L. E., Steitz, T. A., & Simon, M. I. (1987) *Science 235*, 777.

Clark-Lewis, I., Aebersold, R., Ziltener, H., Schrader, J., Hood, L. E., & Kent, S. B. H. (1986) *Science 231*, 134.

Dervan, P. B. (1986) *Science 232*, 464.

Dreyer, J. B., & Dervan, P. B. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 968.

Fried, M., & Crothers, D. (1981) *Nucleic Acids Res. 9*, 6505.

Garner, M. M., & Revzin, A., (1981) *Nucleic Acids Res. 9*, 3047.

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97.

Glasgow, A. C., Bruist, M. F., & Simon, M. I. (1989a) *J. Biol. Chem. 264*, 10072.

Glasgow, A. C., Hughes, K. T., & Simon, M. I. (1989b) in *Mobile DNA* (Berg, D., & Howe, M. M., Eds.) pp 637–659, American Society for Microbiology, Washington, DC.

Griffin, J. H., & Dervan, P. B. (1987) *J. Am. Chem. Soc. 109*, 6840.

Griffin, L. C., & Dervan, P. B. (1989) *Science 245*, 967.

Hertzberg, R. P., & Dervan, P. B. (1982) *J. Am. Chem. Soc. 104*, 313.

Hertzberg, R. P., & Dervan, P. B. (1984) *Biochemistry 23*, 3934.

Johnson, R. C., & Simon, M. I. (1985) *Cell 41*, 781.

Johnson, R. C., Bruist, M. F., Glaccum, M. B., & Simon, M. I. (1984) *Cold Spring Harbor Symp. Quant. Biol. 49*, 751.

Jordan, S. R., & Pabo, C. O. (1988) *Science 242*, 893.

Kent, S. B. H. (1988) *Annu. Rev. Biochem. 57*, 957.

Kent, S. B. H., & Clark-Lewis, I. (1985) in *Synthetic Peptides in Biology and Medicine* (Alitalo, K., Partanen, P., & Vaheri, A. Eds.) pp 29–57, Elsevier, Amsterdam, The Netherlands.

Kent, S. B. H., Hood, L. E., Beilan, H., Meister, S., & Geiser, T. (1984) in *Peptides 1984*, Proceedings of the 18th European Peptide Symposium, Djuroenaeset, Sweden, 1984 (Ragnarsson, U., Ed.) pp 185–188, Almqvist & Wiksell, Stockholm, Sweden.

Kent, S. B. H., Hood, L. E., Beilan, H., Bridgham, J., Marriott, M., Meister, S., & Geiser, T. (1985) in *Peptide Chemistry 1984*, Proceedings of the Japanese Peptide Symposium (Isumiya, N., Ed.) pp 167–170, Protein Research Foundation, Osaka, Japan.

Kent, S. B. H., Parker, K. F., Schiller, D. L., Wood, D. D.-L., Clark-Lewis, I., & Chait, B. T. (1988) in *Peptides: Chemistry and Biology*, Proceedings of the Tenth American Peptide Symposium, St. Louis, MO, 1987 (Marshall, G. R., Ed.) pp 173–178, ESCOM, Leiden, The Netherlands.

Landschulz, W. H., Johnson, P. F., & McKnight, S. L. (1988) *Science 240*, 1759.

Lea, M. S., Gippert, G. P., Saman, R. V., Case, D. A., & Wright, P. (1989) *Science 245*, 635.

Lund, T., Dahl, K. H., Mork, E., Holtlund, J., & Laland, S. G. (1975) *Biochem. Biophys. Res. Commun. 146*, 725.

Mack, D. P., Sluka, J. P., Shin, J. A., Griffin, J. H., Simon, M. I., & Dervan, P. B. (1990) *Biochemistry* (following paper in this issue).

McClarin, J. A., Frederick, C. A., Wang, B. D., Greene, P., Boyer, H. W., Grable, J., & Rosenberg, J. M. (1986) *Science 234*, 1526.

Merrifield, R. B. (1969) *Adv. Enzymol. Relat. Areas Mol. Biol. 32*, 221.

Miller, J., McLachlan, A. D., & Klug, A. (1985) *EMBO J. 4*, 1609.

Moser, H., & Dervan, P. B. (1987) *Science 238*, 645.

Oakley, M. G., & Dervan, P. B. (1990) *Science 248*, 847.

Otwinowski, Z., Schevitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., & Sigler, P. B. (1988) *Nature 335*, 321.

Pabo, C. O., & Sauer, R. T. (1984) *Annu. Rev. Biochem. 53*, 293.

Párraga, G., Horvath, S. J., Eisen, A., Taylor, W. E., Hood, L., Young, E. T., & Klevit, R. E. (1988) *Science 241*, 1489.

Párraga, G., Horvath, S., Hood, L., Young, E. T., & Klevit, R. E. (1990) *Proc. Natl. Acad. Sci. U.S.A.* (in press).
Povsic, T. J., & Dervan, P. B. (1989) *J. Am. Chem. Soc. 111*, 3059.
Sarin, V. K., Kent, S. B. H., Tom, J. P., & Merrifield, R. B. (1981) *Anal. Biochem. 117*, 147.
Schultz, P. G., Taylor, J. S., & Dervan, P. B. (1982) *J. Am Chem. Soc. 104*, 6861.
Sluka, J. P., Horvath, S. J., Bruist, M. F., Simon, M. I., & Dervan, P. B. (1987) *Science 238*, 1129.
Sluka, J. P., Griffin, J., Mack, D. P., & Dervan, P. B. (1990) *J. Am. Chem. Soc.* (in press).
Solomon, M. J., Strauss, F., & Varshavsky, A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1276.
Stewart, J. M., & Young, J. D. (1984) *Solid Phase Peptide Synthesis*, 2nd ed., Pierce Chemical Co., Rockford, IL.
Strobel, S. A., Moser, H. E., & Dervan, P. B. (1988) *J. Am. Chem. Soc. 110*, 7927.
Suzuki, M. (1989) *EMBO J. 8*, 797.
Taylor, J. S., Schultz, P. G., & Dervan, P. B. (1984) *Tetrahedron 40*, 457.
Tullius, T., Dombroski, B., Churchill, M., & Kan, L. (1987) *Methods Enzymol. 155*, 537.
Vinson, C. R., Sigler, P. B., & McKnight, S. L. (1989) *Science 246*, 911.
Wade, W. S., & Dervan, P. B (1987) *J. Am. Chem. Soc. 109*, 1574.
Yang, C. C., & Nash, H. A. (1989) *Cell 57*, 869.
Youngquist, R. S., & Dervan, P. B. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2565.
Zieg, J., Silverman, M., Hilmen, M., & Simon, M. (1977) *Science 196*, 170.

Orientation of the Putative Recognition Helix in the DNA-Binding Domain of Hin Recombinase Complexed with the Hix Site†

David P. Mack,‡ James P. Sluka,‡ Jumi A. Shin,‡ John H. Griffin,‡ Melvin I. Simon,§ and Peter B. Dervan*,‡
Division of Chemistry and Chemical Engineering and Division of Biology, California Institute of Technology, Pasadena, California 91125

Received January 5, 1990; Revised Manuscript Received March 6, 1990

ABSTRACT: On the basis of sequence similarity with other known DNA-binding proteins, the DNA-binding domain of Hin recombinase, residues 139–190, is thought to bind DNA by a helix–turn–helix motif. Two models can be considered that differ in the orientation of the recognition helix in the major groove of DNA. One is based on the orientation of the recognition helix found in the 434 repressor (1–69) and λ repressor–DNA cocrystals, and the other is based on the NMR studies of *lac* repressor headpiece. Cleavage by EDTA·Fe attached to a lysine side chain ($Ser^{183} \rightarrow Lys^{183}$) near the COOH terminus of Hin(139–184) reveals that the putative recognition helix is oriented toward the center of the inverted repeats in a manner similar to that seen in the 434 and λ repressor–DNA cocrystals.

The structural class of DNA-binding proteins best characterized by crystallographic studies contains the helix–turn–helix motif. Comparison of the three dimensional structures of λ cro, λ repressor, and catabolite gene activator protein (CAP) led to the postulate that a conserved α-helix–turn–α–helix motif is involved in recognition of DNA in the major groove and may be a common structural motif for sequence-specific DNA affinity (Anderson et al., 1981; McKay & Steitz, 1981; Pabo & Lewis, 1982; McKay et al., 1982; Ohlendorf & Matthews, 1983; Pabo & Sauer, 1984; Schevitz et al., 1985). The X-ray structure determination of three proteins containing helix–turn–helix motifs bound to their DNA operator sites elucidates the DNA-binding domain of the 434 repressor (1–69) (Anderson et al., 1985, 1987; Aggarwal et al., 1988), the DNA-binding domain of λ repressor (1–92) (Jordan & Pabo, 1988), and the *trp* repressor (Otwinowski et al., 1988). These high-resolution views reveal the complexity of protein–DNA interfaces. The α helices, linked by turns of varying length, contact the sugar–phosphate backbone as well as the base pairs in the major groove. The specificity depends on a set of correlated interactions, and changing any one may affect others (Jordan & Pabo, 1988; Aggarwal et al., 1988). The combination of direct protein–DNA contacts mediated by multiple hydrogen bonds and the sequence-dependent conformational effects in DNA limits our ability to make detailed structural predictions, even if a new protein can be assigned to a structural class such as the helix–turn–helix. Several more examples of protein-DNA complexes containing helix–turn–helix motifs will be needed to determine whether general principles emerge (Otwinowski et al., 1988; Aggerwal et al., 1988; Jordan & Pabo, 1988).

DNA-Binding Domain of Hin Recombinase. Hin recombinase is a 190-residue enzyme thought to recognize DNA by use of a helix–turn–helix motif on the basis of sequence similarities with repressor DNA-binding proteins (Pabo & Sauer, 1984). The 52 residues at the COOH terminus of Hin(139–190) have been shown to contain the DNA-binding domain (Bruist et al., 1987). Affinity cleaving studies using Hin(139–190) equipped with the DNA-cleaving moiety ethylenediaminetetraacetic acid (EDTA)·Fe at the NH_2 terminus have revealed that the NH_2 terminus of the DNA-binding domain of Hin lies proximal to the minor groove near the symmetry axis of Hin recombination sites (Sluka et al., 1987,

† Supported by grants from the DARPA University Research Initiation Program, the National Foundation for Cancer Research, and the NSF and by National Research Service Awards from NIGMS to D.P.M., J.P.S., and J.A.S.
* To whom correspondence should be addressed.
‡ Division of Chemistry and Chemical Engineering.
§ Division of Biology.

0006-2960/90/0429-6561$02.50/0