Site-specific Photo-cross-linking between λ Integrase and Its DNA Recombination Target*

Margaret J. Kovach, Radhakrishna Tirumalaï, and Arthur Landý

From the Department of Molecular Biology, Cell Biology, and Biochemistry, Brown University, Providence, Rhode Island 02912, and Southern Illinois University School of Medicine, Springfield, Illinois 62794-9626, and Celera Genomics, Proteomics Division, Rockville, Maryland 20850

The site-specific recombinase (Int) of bacteriophage λ is a heterobivalent DNA-binding protein and is composed of three domains as follows: an amino-terminal domain that binds with high affinity to “arm-type” sequences within the recombination target DNA (att sites), a carboxyl-terminal domain that contains all of the catalytic functions, and a central domain that contributes significantly to DNA binding at the “core-type” sequences where DNA cleavage and ligation are executed. We constructed a family of core-type DNA oligonucleotides, each of which contained the photoreactive analog 4-thiothymidine (4-thioT) at a different position. When tested for their respective abilities to promote covalent cross-links with Int after irradiation with UV light at 366 nm, one oligonucleotide stood out dramatically. The 4-thioT substitution on the DNA strand opposite the site of Int cleavage led to photo-induced cross-linking efficiencies of ~20%. The efficiency and specificity of Int binding and cleavage at this 4-thioT-substituted core site was shown to be largely uncompromised, and its ability to participate in a full site-specific recombination reaction was reduced only slightly. Identification of the photo-cross-linked residue as Lys-141 in the central domain provides, along with other results, several insights about the nature of core-type DNA recognition by the bivalent recombinases of the λ Int family.

Each att site contains an inverted pair of “core-type” Int-binding sites (9 bp each) with a centered 7-bp “overlap region.” A reciprocal exchange of the “top” strands at the left boundary of the overlap region generates a Holliday junction recombination intermediate, which is then resolved by exchange of the “bottom” strands at the right boundary of the overlap region. DNA cleavage is mediated by a tyrosine hydroxyl that attacks the scissile phosphate, forming a 3′-phosphotyrosyl link to the nicked DNA. This covalent protein-DNA intermediate is resolved when the 5′-terminal hydroxyl of the invading DNA strand attacks the phosphotyrosyl linkage and displaces the protein. The chemistry of DNA strand exchange and the general arrangement of the “core region” are common to all of the λ Int family members except that their overlap regions vary from 6 to 8 bp in length, and their core-type binding sites vary from 9 to 13 bp. For some family members, such as Cre, XerC/D, and FLP, the core region composes the entire (minimal) att site. For other family members, such as λ and HP1 integrases, the att sites are more complex and contain additional protein-binding sites in viral DNA sequences that compose flanking “arms.” Some of these flanking sites bind to DNA bending accessory proteins like IHF, Xis, and Fis, whereas others bind to the amino-terminal domain of Int, resulting in a higher order complex where Int bridges the core and arm sequences of a sharply bent att DNA.

Underlying the various layers of complexity that divide λ Int family members into several subgroups is a common carboxyl-terminal catalytic domain that executes the cleavage and rejoining of core-type DNA sequences in an energy-conserving reaction pathway that is the hallmark of these recombinases. The carboxyl-terminal “domain” of λ Int (residues 65–356) binds with low affinity to core-type sites located at the positions of strand cleavage and functions as a topoisomerase. This C65 can be further dissected by proteolysis into two smaller domains, encompassing residues 65–169 and 170–356, respectively. The latter, termed C170 or the catalytic domain, has been characterized and its crystal structure has been determined (6, 7).

The C170 domain contains all of the residues that have been identified as being conserved in the λ Int family of recombinases (8–11). These include a very highly conserved triad of Arg-212, His-308, and Arg-311 that has been suggested to activate the scissile phosphate for DNA cleavage (12, 13), the active site nucleophile Tyr-342 (14), and Glu-174, which can be mutated to give a hyper-recombination phenotype (15). The C170 domain of λ Int is approximately the same size as the smallest λ Int family members, such as FimB and FimE of E. coli (227 and 209 amino acids respectively) (16). Crystal structures of the catalytic domain of HP1 integrase (17), the XerD recombinase (18), and the Cre recombinase complexed with its att site loxP (19) reveal protein folds resembling that of the λ Int domain.
Finally, the Cre/lox binding to terminal (catalytic) and amino-terminal (arm-binding) domains because it is flanked by the carboxyl-specific DNA binding affinity and sequence recognition specificity of Int C65Y (25 pmol) and allowed to proceed for 1 h. The reactions were cooled/annealing period of 4°C.

Int C170 domain (7). However, there are significant differences in the active sites of these recombinases, including different orientations of the tyrosine nucleophile and differences in nearby segments that are predicted to contact the DNA.

In all of the λ Int family members there is a second domain just upstream of the catalytic domain that is also involved in binding to core-type sites. In λ Int this is called the central core binding (CB) domain because it is flanked by the carboxyterminal (catalytic) and amino-terminal (arm-binding) domains. In the monovalent λ Int family members that bind only to core-type sites (e.g. Cre, Flp, and XerCD), the analogous domain is an amino-terminal domain.

Previous studies in our laboratory (6) had shown that although the catalytic domain of λ Int catalyzes the cleavage and rejoining of core-type DNA sequences, it does not form electrophoretically stable complexes with atr DNA. Experiments specifically designed to identify the interface(s) between λ Int and core-type DNA all pointed to the CB domain as follows: zero length UV-induced photo-cross-linking identified Ala-125 and Ala-126, DNA-sensitive modification by pyridoxal 5'-phosphate identified Lys-103, and an isolated CB domain formed stable complexes with core-type DNA (20). These results were unexpected because a number of experiments from other laboratories all pointed to the primary of the catalytic domain in binding to core-type DNA. Additionally, several mutations affecting core-type DNA binding affinity and sequence recognition specificity were also found to be located in the catalytic domain (21, 22).

The catalytic domains of Cre and FLP can bind autonomously to core-type sites, i.e. in the absence of their amino-terminal domains, which correspond to the CB domain of λ Int (23, 24).

The Cre/loxA co-crystal structure indicated that upon binding to lox (core-type) DNA, the catalytic domain buries ~50% more solvent-accessible surface area at the DNA interface than the amino-terminal (CB-analogous) domain (19).

To explore further these apparent dichotomies, we looked for a different chemistry with which to probe the Int-core DNA interface(s). We found that incorporation of the photoaffinity cross-linking analog, 4-thio-deoxythymidine (4-thioT), at a unique position in the core DNA yields high efficiency photo-cross-linking to Int. Identification of the photo-cross-linked residue as Lys-141 in the CB domain provides, along with other results, several insights about the nature of core-type DNA recognition by the bivalent recombinases of the λ Int family. An additional bonus of these experiments is the very high efficiency with which the photo-cross-link is generated, thus providing a potentially useful handle for dissecting the higher order organization of the large complex structures associated with this site-specific recombination pathway.

**EXPERIMENTAL PROCEDURES**

**Protein Expression and Purification**—Full-length bacteriophage λ Int, Int C65Y, and Int C65F proteins were produced in *E. coli* strain BL21 from expression plasmids under the control of bacteriophage T7 promoter and were purified from the insoluble (λ Int) and soluble (C65Y and C65F forms) fractions of the cell lysate (20). IIF (25, 26) and Xis (27, 28) were also purified from overproducing strains. The protein preparations, purified by a series of phosphocellulose (Whatman), SP-Sepharose (Amersham Biosciences), and hydroxyapatite (Calbiochem) chromatography, were estimated to be purified to greater than 95% pure as evaluated by Coomassie staining of an overloaded SDS-polyacrylamide gel. Protein concentrations were estimated by the dye-binding method (29).

**Oligonucleotides**—Synthetic oligonucleotides were purchased from Cruachem, Inc. (Dulles, VA), in lyophilized form. Before use, the oligonucleotides were resuspended in TE buffer at 1 mM. The sequences for each 4-thioT-substituted core-binding site (underlined) are as follows (where ts designates top strand, bs designates bottom strand).

**Photo-Cross-linking of Int C65F to 4-ThioT-substituted Half-att Site**

**FIG. 1. Schematic illustration of Int-dependent recombination and the ensemble of relevant protein-binding sites.** Integrative recombination between the phage (attP) and bacterial (attB) att sites (bottom to top) requires the phage-encoded Int protein and the host-encoded accessory protein IHF and generates the left (attL) and right (attR) prophage att sites. Excessive recombination between attL and attR (top to bottom) additionally requires the phage-encoded Xis protein and is stimulated by the host-encoded Fis protein. Int has five high affinity arm-type-binding sites (circles) and four lower affinity core-type sites that flank the 7-bp overlap region (solid bar) as inverted repeats (opposing open arrows) and encompass each of the DNA cleavage sites (curved solid arrows). The sequence of the C’ core-type site (used in this study) and an abutting portion of the overlap region are shown in the box along with the location of the bottom strand cleavage site. IHF has three binding sites (inverted triangles), and Xis has two binding sites (upright triangles), one of which overlaps the single Fis site (F).

Int C65F was produced in *E. coli* strain BL21 from expression plasmids under the control of bacteriophage T7 promoter and was purified from the insoluble (λ Int) and soluble (C65Y and C65F forms) fractions of the cell lysate (20). IIF (25, 26) and Xis (27, 28) were also purified from overproducing strains. The protein preparations, purified by a series of phosphocellulose (Whatman), SP-Sepharose (Amersham Biosciences), and hydroxyapatite (Calbiochem) chromatography, were estimated to be pure to greater than 95% pure as evaluated by Coomassie staining of an overloaded SDS-polyacrylamide gel. Protein concentrations were estimated by the dye-binding method (29).

**Oligonucleotides**—Synthetic oligonucleotides were purchased from Cruachem, Inc. (Dulles, VA), in lyophilized form. Before use, the oligonucleotides were resuspended in TE buffer at 1 mM. The sequences for each 4-thioT-substituted core-binding site (underlined) are as follows (where ts designates top strand, bs designates bottom strand).

**Photo-Cross-linking of Int C65F to 4-ThioT-substituted Half-att Site**

**FIG. 1. Schematic illustration of Int-dependent recombination and the ensemble of relevant protein-binding sites.** Integrative recombination between the phage (attP) and bacterial (attB) att sites (bottom to top) requires the phage-encoded Int protein and the host-encoded accessory protein IHF and generates the left (attL) and right (attR) prophage att sites. Excessive recombination between attL and attR (top to bottom) additionally requires the phage-encoded Xis protein and is stimulated by the host-encoded Fis protein. Int has five high affinity arm-type-binding sites (circles) and four lower affinity core-type sites that flank the 7-bp overlap region (solid bar) as inverted repeats (opposing open arrows) and encompass each of the DNA cleavage sites (curved solid arrows). The sequence of the C’ core-type site (used in this study) and an abutting portion of the overlap region are shown in the box along with the location of the bottom strand cleavage site. IHF has three binding sites (inverted triangles), and Xis has two binding sites (upright triangles), one of which overlaps the single Fis site (F).

Int C170 domain (7). However, there are significant differences in the active sites of these recombinases, including different orientations of the tyrosine nucleophile and differences in nearby segments that are predicted to contact the DNA.

In all of the λ Int family members there is a second domain just upstream of the catalytic domain that is also involved in binding to core-type sites. In λ Int this is called the central core binding (CB) domain because it is flanked by the carboxyterminal (catalytic) and amino-terminal (arm-binding) domains. In the monovalent λ Int family members that bind only to core-type sites (e.g. Cre, Flp, and XerCD), the analogous domain is an amino-terminal domain.

Previous studies in our laboratory (6) had shown that although the catalytic domain of λ Int catalyzes the cleavage and rejoining of core-type DNA sequences, it does not form electrophoretically stable complexes with atr DNA. Experiments specifically designed to identify the interface(s) between λ Int and core-type DNA all pointed to the CB domain as follows: zero length UV-induced photo-cross-linking identified Ala-125 and Ala-126, DNA-sensitive modification by pyridoxal 5'-phosphate identified Lys-103, and an isolated CB domain formed stable complexes with core-type DNA (20). These results were unexpected because a number of experiments from other laboratories all pointed to the primary of the catalytic domain in binding to core-type DNA. Additionally, several mutations affecting core-type DNA binding affinity and sequence recognition specificity were also found to be located in the catalytic domain (21, 22).

The catalytic domains of Cre and FLP can bind autonomously to core-type sites, i.e. in the absence of their amino-terminal domains, which correspond to the CB domain of λ Int (23, 24). Finally, the Cre/loxA co-crystal structure indicated that upon binding to lox (core-type) DNA, the catalytic domain buries ~50% more solvent-accessible surface area at the DNA interface than the amino-terminal (CB-analogous) domain (19).

To explore further these apparent dichotomies, we looked for a different chemistry with which to probe the Int-core DNA interface(s). We found that incorporation of the photoaffinity cross-linking analog, 4-thio-deoxythymidine (4-thioT), at a unique position in the core DNA yields high efficiency photo-cross-linking to Int. Identification of the photo-cross-linked residue as Lys-141 in the CB domain provides, along with other results, several insights about the nature of core-type DNA recognition by the bivalent recombinases of the λ Int family. An additional bonus of these experiments is the very high efficiency with which the photo-cross-link is generated, thus providing a potentially useful handle for dissecting the higher order organization of the large complex structures associated with this site-specific recombination pathway.

**EXPERIMENTAL PROCEDURES**

**Protein Expression and Purification**—Full-length bacteriophage λ Int, Int C65Y, and Int C65F proteins were produced in *E. coli* strain BL21 from expression plasmids under the control of bacteriophage T7 promoter and were purified from the insoluble (λ Int) and soluble (C65Y and C65F forms) fractions of the cell lysate (20). IIF (25, 26) and Xis (27, 28) were also purified from overproducing strains. The protein preparations, purified by a series of phosphocellulose (Whatman), SP-Sepharose (Amersham Biosciences), and hydroxyapatite (Calbiochem) chromatography, were estimated to be pure to greater than 95% pure as evaluated by Coomassie staining of an overloaded SDS-polyacrylamide gel. Protein concentrations were estimated by the dye-binding method (29).

**Oligonucleotides**—Synthetic oligonucleotides were purchased from Cruachem, Inc. (Dulles, VA), in lyophilized form. Before use, the oligonucleotides were resuspended in TE buffer at 1 mM. The sequences for each 4-thioT-substituted core-binding site (underlined) are as follows (where ts designates top strand, bs designates bottom strand).
Substrates—Binding of Int C65F (25 pmol) to radiolabeled half-att site DNA substrates modified with 4-thioT substitutions proceeded at 25 °C for 20 min in a 10-µl mixture of 10 mM Tris-HCl (pH 7.5), 50 mM NaCl, 5% glycerol, 1 mM EDTA, 1 mg/ml bovine serum albumin, and 1 µg/ml sheared salmon sperm DNA. The samples were then transferred to the surface of the polyacrylamide gel plate pre-chilled to 4 °C. Photo-cross-linking of the samples was achieved by exposing the samples to a 366 nm UV light source (Blak-Ray lamp, model UVL-56, Ultra-Violet products, Inc., San Gabriel, CA) positioned 2 cm away from the samples. Photo-cross-linking was performed on ice at 4 °C for a period of 45 min. Non-covalent interactions were disassociated by addition of 0.2% SDS. Photo-cross-linking of an Int C65F titration was also performed as described above, except that a constant amount of radiolabeled T3 oligomer (25 pmol) was mixed with Int C65F (500, 250, 125, 60, 30, 15, and 7.5 pmol) prior to cross-linking. UV-dependent covalent complexes were separated by electrophoresis through 10% polyacrylamide gel in 0.5 7.5 pmol) prior to cross-linking. UV-dependent covalent complexes were separated by electrophoresis through 10% polyacrylamide gel in 0.5× TBE, 0.1% SDS buffer. The gels were exposed on a PhosphorImager plate and also autoradiographed on Fuji x-ray film. The PhosphorImager plate was scanned with a Fuji BAS 1000 scanner, and the image was quantitated with the Fuji MacBAS software package.

Gel Mobility Shift Assays of Int C65F Binding to T5–4-ThioT-substituted and Non-substituted Half-att Sites—Several concentrations of Int C65F (5, 6, 12, 25, and 50 pmol) were mixed with radiolabeled core-type oligomers (25 pmol) and incubated at 25 °C for a 10 µl mixture of 10 mM Tris-HCl (pH 7.5), 50 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol, and 5% glycerol. The reactions were analyzed by electrophoresis through an 8% polyacrylamide gel in 25 mM Tris, 200 mM glycine, 1 mM EDTA buffer. The gels were exposed on a PhosphorImager plate and also autoradiographed on Fuji x-ray film. The PhosphorImager plate was scanned with a Fuji BAS 1000 scanner, and the image was visualized and quantitated with the Fuji MacBAS software package.

Competition binding assays were set up as described above, except that Int C65F (12.5 pmol) was incubated with radiolabeled T3 4-thioT duplex (25 pmol) in the presence of increasing amounts of non-labeled competitor DNA (25, 50, 75, 100, and 150 pmol). The homologous competitor shared the same nucleotide sequence as the T3 4-thioT duplex except that it did not contain the 4-thioT substitution. The heterogeneous competitor shared the same base composition as the T3 duplex, but the sequence for the core binding domain was scrambled (see above). The gels were quantitated on the PhosphorImager.

Excissive Recombination Assays—attL and attR plasmids containing the overlap sequences 5′-TTTATAX (pPEG1) and 5′-TTTATAT (pPEG2), respectively, were generated to replicate the overlap sequence of the half-att site. (Note that a bold X indicates the 4-thioT substitution.) 4-ThioT-substituted and non-substituted attL substrates were generated by PCR amplification using the appropriate attL primer along with a downstream primer homologous to a sequence within the pBR327 backbone of the plasmid pPEG1 (see under “Oligonucleotides”). The attL substrates (~400 bp) were gel-purified and 5′-labeled using T4 polynucleotide kinase with 32P-ATP. Unincorporated 32P-ATP was removed via a Sephadex G-25 spin column (Pharmacia). Recombination competition assays were carried out at 25 °C for 4 h in a 20-µl mixture containing 0.10 pmol of supercoiled attR DNA substrate and 0.05 pmol of radiolabeled linear attL substrate (4-thioT-modified or non-modified), 25 mM MOPS (pH 7.9), 50 mM NaCl, 5 mM EDTA, 6 mM spermidine, 2.5 mM dithiothreitol, and 0.5 µg/ml bovine serum albumin. Reactions were preincubated with 1.5 units of IHF for 20 min before the addition of 1.25 pmol of X. The reaction mixtures were then diluted to 2 ml urea by the addition of 3 volumes of 100 mM NH4HCO3, to make the conditions compatible with trypsin digestion. The digestion was allowed to go to completion (~24 h), and the Int C65F photo-cross-linked to T3 4-thioT-att DNA was recovered by anion-exchange HPLC on a Synchron AX 300 (250 × 4.6 mm) column, equilibrated with 20 mM sodium phosphate (pH 6.8), 20% (v/v) acetonitrile (equilibration buffer). The peptides were eluted by increasing the concentration of NaCl in the equilibration buffer as follows: 0–30 min, no salt; 30–90 min, a gradient from 0 to 1.0 M NaCl; 90–110 min, 1 M NaCl. The flow rate was 1.0 ml/min. All HPLC analyses were conducted using a Varian 920 inert solvent delivery system equipped with a polychrome 90065 diode array detector. peptide peaks were monitored at 254 nm. Fractions 71–76 were pooled and further purified by reverse phase HPLC on a C18 column (Vydac) under ion-pairing conditions. The column was equilibrated with 10 mM triethylammonium acetate (pH 7.0) for 10 min. After a 5-min wash with HPLC-grade water, the peptide(s) were eluted by increasing the concentration of acetonitrile in water from 10% to 30%. The peptides were monitored at 254 nm. Fractions corresponding to peaks I (residues 67–82) and II (residues 90–102) were pooled individually and subjected to amino-terminal amino acid sequencing using the Edman degradation reaction (30) on an Applied Biosystems 470A gas phase sequencer at the W. M. Keck Foundation Biotechnology Resource Laboratory, Yale University, New Haven, CT. The resulting phenylthiohydantoin derivatives were analyzed using an on-line Applied Biosystems model 470A microbore HPLC.

RESULTS

Scanning Core Sites with 4-ThioT Substitutions—A Int and its close relatives are unusual in being heterodivalent DNA-binding proteins. A short amino-terminal domain, residues 1–64, is responsible for binding to “arm-type” DNA sequences distant from the core-type DNA sequences where DNA strands are cleaved, exchanged, and religated. All of the operations on core-type DNA, including binding, are governed by the carboxyl-terminal portion of Int (residues 65–356). This autonomously functioning protein, called C65, has been cloned and purified (20) and, unless noted otherwise, is considered exclusively in this paper. Its activity on core-type DNA is not only undiminished but is actually enhanced by separation from the amino-terminal domain (31). To monitor the efficiency of Int C65 binding and DNA cleavage, we took advantage of a suicide core site (32). As diagrammed in Fig. 2, these substrates contain a 3′-thymine three bases from the scissile phosphate. When Int cleaves this substrate (via formation of a covalent 3′-phosphotyrosine linkage) it generates a 3-base oligonucleotide that diffuses away from the att site DNA. Loss of the oligonucleotide removes the 5′-OH nucleophile that would otherwise attack the phosphotyrosine bond to reform the phosphodiester DNA linkage and release Int. In the absence of the 5′-OH nucleophile, the phosphotyrosine linkage is stable, and the covalent complex is readily monitored or isolated by gel electrophoresis in SDS-polyacrylamide.

In the experiments reported here we used short synthetic oligonucleotides (22/23- or 30/34-mers) containing a “half-att” site. This consists of a single 8-bp core-type Int-binding site and a portion of the 7-bp overlap region. The cleaved strands (top strands in all of the figures) contained 3 bases of the overlap region, which are lost upon Int cleavage. The bottom strands contained either 4 or the full 7 bases of the overlap region. We have not observed any differences in binding or cleavage efficiency between substrates with 4–7 bases in the bottom strand of the overlap region nor between substrates containing 12 versus 20 bp preceding the Int-binding site (data not shown).

We constructed 11 half-att site substrates, each containing a 4-thioT substitution at a different position and labeled with 32P at the 5′ terminus of the top strands. Because we are incorporating a thio analog, and in some cases substituting a thymine for the canonical base, we tested each substrate for its ability to be cleaved by wild-type Int and also as a substrate for the covalent complex (as diagrammed in Fig. 2) which is seen as a band with slower electrophoretic mobility in SDS-polyacrylamide than the free DNA. Most of the substrates were very efficient at forming covalent complex (Fig. 3A). (A lane showing cleavage of the unsubstituted att site was not included in the gel shown in Fig. 3 but in similar experiments its efficiency is approximately the...
cross-linking is approximately equal to that of the unsubstituted core-type site (0.5%) (data not shown). Whereas these two substitutions may be poor photo-cross-linking substrates because the 4-thioTs intrude too severely into the Int-binding space (to the point of preventing Int cleavage), there will be other substitutions that are poor photo-cross-linking substrates because the thio group is too far from the bound Int (or from a reactive residue within Int). In this category of att sites, which are cleaved efficiently by Int but not efficiently photo-cross-linked (<3%), we find the B6, T8, T9, B11, and B12 substitutions. The group of att sites yielding intermediate photo-cross-linking efficiencies (~5%) consists of the B4, T11, and T12 substitutions. The T3 substitution, which consistently yielded photo-cross-linking efficiencies of ~20%, was clearly the best of those we tested. It should be noted that the level of cross-linking of even the poorest 4-thioT substrates is still well above the 0.5% that is obtained with unsubstituted att sites (data not shown).

**Functionality of the T3 4-ThioT att Site**—The high efficiency of photo-cross-linking with the T3 att site suggested that this might be useful for identifying the specific target residues in λ Int. However, it was first necessary to establish that the 4-thioT substitution at this position does not substantially impair its interaction with Int. It was shown in Fig. 3A that the T3 substitution did not interfere with Int cleavage. Nevertheless, it should be pointed out that although binding is obviously a precursor for cleavage, the latter results in the formation of a stable covalent product whose accumulation might mask a reduction in binding, especially if cleavage is rate-limiting. We therefore also tested this substrate for Int binding in studies that were carried out with the mutant C65 Int lacking the tyrosine 342 nucleophile (Y342F). As seen in Fig. 4A the binding efficiency of the 4-thioT-substituted half-att site is almost the same as that of the unsubstituted half-att site. The specificity of binding to the substituted half-att site was assayed by its relative sensitivity to competition from a heterologous versus homologous competitor. The homologous competitor half-att site has the same sequence as the labeled 4-thioT-substituted half-att site except that it has a thymine in place of the 4-thioT. The heterologous competitor had the same base composition as the 4-thioT substrate, but the core-binding site was scrambled (see “Experimental Procedures”). As seen in Fig. 4B, Int binding to the 4-thioT-substituted half-att site was considerably more resistant to competition from the heterologous competitor. The observed low level of competition by heterologous competitor is consistent with previous experiments showing that Int binding to single DNA sites is not highly specific (33). The approximately 5-fold difference in competition by heterologous and homologous competitors is consistent with specific binding of Int to the 4-thioT-substituted att site. An additional demonstration of this specificity is provided by the fact that the 4-thioT-substituted and unsubstituted half-att sites are equally efficient in competing for binding to a radiolabeled unsubstituted half-att site (data not shown).

As a final test of the acceptability of the 4-thioT substitution, we tested its effect on a full recombination reaction. Because the thymine for cytosine substitution at position 3 is within the overlap region, and sequence identity is required between the recombination partners, it was necessary to construct both attL and attR DNAs containing the desired substitution. To do this it was first necessary to construct plasmids containing the appropriate attL and attR sites, pPEG1 and pPEG2, respectively (see “Experimental Procedures”). The attL plasmid DNA was used as a template for generating 4-thioT-substituted and non-substituted attL substrates by PCR amplification reactions. One primer, with or without the 4-thioT modification,
was complementary to the overlap region and directed DNA synthesis into the attL region, whereas the opposing primer annealed to the region derived from the pBR327 vector outside of the attL sequence. The attL PCR products were gel-purified and labeled with 32P at their 5′ termini and were each used in an excisive recombination reaction with the attR isolated from pPEG2. The extent of recombination was determined by electrophoresis on agarose gels and quantitation with a PhosphorImager (Fig. 4C). Recombination of the 4-thioT-substituted attL was ~30% of the unsubstituted attL but was still quite good. There are several steps in the recombination reaction after Int binding and cleavage that might account for the sensitivity to the 4-thioT substitution, but we have not explored this question further. Taken together, the results on Int binding, cleavage, and recombination strongly support the validity of using the 4-thioT substitution at position T3 for studying Int-core site interactions.

Identification of the Photo-cross-linked Residue in Int—In preparation for preparative scale photo-cross-linking, we determined that the optimal molar ratio of Int to 4-thioT-substituted att site DNA was ~1:1 (Fig. 5). We also modified a number of reaction conditions such as salt, glycerol, and buffer concentrations for their effect on the efficiency of cross-linking before arriving at the conditions described under “Experimental Procedures” (data not shown). In a typical preparative scale photo-cross-linking reaction Int C65F (150 nmol) was incubated with the 4-thioT-substituted att site DNA (125 nmol) in 1 ml of binding buffer for 20 min at 25 °C. The reaction was then irradiated at 4 °C for 45 min as described under “Experimental Procedures”. The reaction mixture was brought to 8 M urea, and the protein plus cross-linked DNA were separated from most (~95%) of the non-cross-linked DNA by repeated cycles of filtration and washing in a Centricon-30 concentrator. The denatured protein and protein-DNA complex was brought to 2 M urea before being digested to completion with trypsin (~24 h). The resultant peptides were separated by anion-exchange HPLC, which is a particularly effective purification of the peptides that are cross-linked to DNA because Int is such a highly basic protein (Fig. 6A). We established that non-cross-linked peptides and non-cross-linked Int protein do not bind to the column under these conditions (data not shown). Fractions 71–78 were pooled for further purification by reverse phase HPLC on a C18 column. From the reverse phase column (Fig. 6B) fractions corresponding to peaks I and II were pooled, lyophilized, and prepared for amino-terminal sequencing. Fractions from peak II did not yield any amino acid sequence and

![Figure 3](https://example.com/figure3.png)

**Fig. 3.** Formation of protein-DNA covalent complexes through enzymatic cleavage or 4-thioT photo-cross-linking. **A**, radiolabeled half-att site suicide substrates with a single 4-thioT:A base pair substitution at the specified positions within and surrounding the core-type recognition sequence were either incubated with Int C65Y (Y) or incubated with Int C65F (F) and irradiated with 366 nm UV light, as described in Fig. 2 and under “Experimental Procedures.” The oligonucleotides are labeled according to the position of the 4-thioT:A base pair substitution (T and B refer to top and bottom strands, respectively) (see B). Samples were assayed for covalent complex formation by gel electrophoresis through 10% polyacrylamide containing 0.1% SDS, and 32P-labeled complexes were visualized by autoradiography. Arrows at the left indicate positions of C65-DNA complexes and free DNA substrates. B, the amount of radioactivity in each band was quantitated using a PhosphorImager, and the bar graph shows the percent of total DNA that is present in the 4-thioT covalent complex. The inset of the wild-type C core-binding site (boxed) shows the canonical base (circled) that has been replaced by the 4-thioT along with the corresponding coordinate for the top (T) or bottom (B) strands. The Int cleavage site is denoted by a curved arrow.
are thought to be free DNA. This is consistent with its elution position, which coincides with the elution position of free DNA alone (data not shown).

Gas phase amino-terminal amino acid sequencing of the final pooled peak I fractions was used to identify the peptide that had been photo-cross-linked to the 4-thioT-substituted DNA. If the photo-induced cross-link is stable under the conditions employed for gas phase sequencing, the cross-linked residue will not be extracted from the Polybrene-coated support disc and consequently there will be a gap at that position in the sequence (34). Eight cycles of sequencing uniquely identified the trypsin cleavage product Ala-Ala-Ser-Ala-Lys-Leu-Ile-Arg, which corresponds to residues 137–144 of λ Int. All of the residues were obtained in good yield except for Lys-141 which was 10% of the expected yield. The yield of Lys-141 in cycle 5 was 10% the yield of Leu-142 in cycle 6 and 25% of Arg-144 in cycle 8 (which would have experienced some washout, because it is the last residue).

We did not obtain clean peptide analyses in experiments where trypsin was replaced by V8 protease either under conditions that favor cleavage after Glu and Asp or only after Glu (35). However, we have considerable confidence in the results obtained with trypsin which were clean, robust, and reproducible (including the 10% yield of Lys-141). The reproducibly low yield of Lys-141 in the cross-linked peptide and the failure of trypsin to cleave after this Lys strongly suggest that is responsible for the photo-induced cross-link with the 4-thioT at position T3 of the attL site.

DISCUSSION

4-Thiopyrimidines have proven to be useful photo-cross-linking reagents in a number of investigations of protein-nucleic acid interactions (36–44). The structure of 4-thioT is very the same base composition as T3, but the sequence is scrambled so that the core-binding site is destroyed. C, excisive recombination efficiencies of T3 4-thioT and non-substituted attLs were assayed at 25 °C for 4 h as described under "Experimental Procedures." Recombination products were analyzed by electrophoresis through a 1.2% agarose gel followed by autoradiography. The level of recombination was quantitated using a PhosphorImager.
similar to that of thymidine because the sulfur atom that replaces the 4-keto oxygen has a Van der Waals radius only 0.45 Å larger than oxygen. This small variation in structure, the unaltered base pairing properties from thymidine, and its high photoreactivity with a wide range of amino acid residues makes 4-thioT an attractive candidate chromophore for studying protein-DNA interactions. Depending upon the protein-DNA complex being studied and the position of the 4-thioT substitution within the DNA photo-cross-linking, efficiencies as high as 35% have been obtained (39).

An important concern in all affinity labeling experiments is whether the observed cross-linking is to residues involved in, or close to, protein-DNA contacts that define specific and mechanistically relevant complexes. There are several arguments supporting the functional relevance of the Lys-141 residue identified here. Specificity at the level of the DNA is suggested by the fact that 4-thioT at different positions within the core-type binding site cross-linked with a range of different efficiencies, as would be expected for protein binding that is unique and specific. The fact that Int binds without severe distortion to a core site with the 4-thioT substitution at position T3 is supported by a binding curve that is quite similar to that of the unsubstituted site (Fig. 4A). The specificity of Int binding to the 4-thioT core site is further supported by the approximately 5-fold difference in sensitivity to competition by heterologous versus homologous competitor DNAs (Fig. 4B) and by the fact that the 4-thioT-substituted and unsubstituted half-att sites are equally efficient in competing for binding to an unsubstituted site (data not shown). The most demanding assay for tolerance of the 4-thioT substitution was its effect on recombination, which was reduced to ~30% of the unsubstituted attL partner. We regard this as an acceptable level of recombination, especially because there are several steps subsequent to binding and cleavage that might also contribute sensitivity to 4-thioT substitution at this position. Indeed, some effect on recombination is consistent with this relatively benign substitution being at, or near, an important protein-DNA interface.

It is interesting to note that the T3 location of the 4-thioT substitutions, is at one of the 2 bp flanking the scissile phosphate (see Fig. 1), neither of which is critical for Int recognition (33); we had interpreted this degeneracy as reflecting the need for flexibility at the site of DNA cleavage and strand exchange. The 4-thioT substitution is in the strand that is not cleaved and therefore also not transferred during recombination to form the heteroduplex overlap region. We, of course, do not know if (any of) these factors are related to the uniquely high efficiency of cross-linking at this position, but it is a very welcome tool. We believe that the 20% cross-linking efficiency is high enough to enable a detailed biochemical analysis of the higher order complexes responsible for a site-specific recombination. Although laser irradiation may further increase the yields of photo-cross-linked product (45), the present efficiencies are already sufficient for a multistep analysis of higher order complexes.

The Lys-141 interaction identified in the present experi-

![Fig. 6. Isolation and purification of Int C65 tryptic peptides photo-cross-linked to T3 4-thioT-att site DNA. Int C65F was photo-cross-linked to the T3 4-thioT half-att site DNA and subjected to trypsin digestion as described under “Experimental Procedures.” The peptides were subjected to anion-exchange HPLC purification (top panel) followed by reverse phase HPLC on a C-18 matrix (bottom panel). The absorbance at 254 nm is plotted as a function of retention time. Fractions 71–78, corresponding to the major peak generated on the anion-exchange HPLC profile, were pooled and subjected to further purification by reverse phase HPLC on a C-18 matrix. Peaks I (fractions 67–82) and II (fractions 90–102) of the C-18 profile were collected and lyophilized in preparation for amino acid sequencing. Fractions from peak II did not yield any amino acid sequence and are thought to be free DNA.](image-url)
ments is not far on the linear Int sequence from two other CB residues we had identified previously as being at or near the interface with core-type DNA (Fig. 7). Ala-125 and Ala-126 were identified by zero length UV cross-linking, and Lys-103 was identified as a residue that reacts with pyridoxal 5′-phosphate in a manner that is sensitive to competition by core-type DNA (20). CB domain residues implicated in core-type DNA recognition by genetic approaches include position 99 in both λ and HK022 Ints and Thr-146 and Asp-149 in HK022 Int (21, 46). We are unable to relate Lys-141, or any of the other CB residues implicated in DNA binding, to the analogous domains in the cocrystal structures of Cre/lox (19) and Flp/FRT (47) because there is so little sequence similarity to either of them (and no structural similarity between them) in this domain.

As noted in the Introduction, we were somewhat surprised by the apparent difference between our biochemical results that pointed to the CB domain as being dominant in core-type DNA binding and other experiments that pointed to the catalytic domain as being more prominent. However, the 4-thioT cross-linking reported here and the other two chemistries used previously (20), all point to the CB domain and not to the catalytic domain. The results of these three very different biochemical approaches, the capacity of the CB domain for autonomous core-type DNA binding and specificity, and the extremely weak binding of the isolated catalytic domain all suggest that in λ Int the CB domain is the major determinant of core-type DNA binding. This is also likely to be the case for the closely related bivalent integrase HP1, whose crystal structure has also been solved, because no DNA binding (or catalytic activity) could be detected for its isolated catalytic domain (17).

Although architectural variations among λ Int family members are not surprising, and even expected because of the very weak sequence homologies, one does have to address the apparently different views obtained from the biochemical versus genetic analyses of core-type specificity in λ Int (21). We suggest that the overall specificity for the core region during recombination is the sum of two kinds of discrimination by λ Int. One is the DNA binding energy and specificity that is contributed primarily (but not entirely) by the CB domain, and the other is the discrimination and specificity that is manifested during catalysis by the catalytic domain. One closely related example of the latter is seen with restriction enzyme EcoRV, which has the same equilibrium constant for binding to specific versus nonspecific DNA but a million-fold difference in cleavage rates (48).

In support of this view we cite recent experiments from Gardner and co-workers (46) that attempted to identify mutations in HK022 Int that would confer the ability to recognize λ core-type DNA. These experiments, which exploited the powerful P22 challenge phage system, measure in vivo DNA binding affinities and therefore differ from the earlier genetic experiments on specificity determinants that utilized a recombination assay to probe DNA recognition specificity. Consistent with our suggestion, the P22 challenge system did not identify any of the four residues in the catalytic domain that were found to confer specificity in recombination, but they did pick up the one previously identified residue in the CB domain as well as two additional residues in the CB domain that increase the general affinity for DNA without altering the discrimination between λ and HK022 core-type DNA sites (46).

Based on the results presented here and from other laboratories, and the considerations discussed above, we suggest that much (but not all) of the binding energy and specificity for interaction with core-type DNA by λ Int and other heterobivalent recombinases comes from the CB domain and that additional discrimination is provided by the catalytic domain during the catalysis of DNA cleavage and ligation.

Acknowledgments—We thank Marco Azaro for comments on the manuscript and help with the figures; Jeffrey Liu and Tina Oliveira for technical assistance; Gregory Van Duyne and members of our laboratory for helpful conversations; and Joan Boyles for preparation of the manuscript and figures.

REFERENCES
1. Kikuchi, Y., and Nash, H. A. (1978) J. Biol. Chem. 253, 7149–7157
2. Nash, H. A. (1996) in Escherichia coli and Salmonella (Neidhardt, F. C., Curtiss, R., III, Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B.,
Site-specific Photo-cross-linking of \( \lambda \) Int

Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., eds) 
Escherichia coli and Salmonella, Cellular and Molecular Biology, pp. 2363–2376, American Society for Microbiology, Washington, D. C.

Hallet, B., and Sherratt, D. J. (1997) FEMS Microbiol. Lett. 21, 157–178

Grainge, I., and Jayaram, M. (1999) Mol. Microbiol. 33, 449–456

Campbell, A. M. (1962) in Advances in Genetics (Caspari, E. W., and Thoday, J. M., eds), pp. 101–145, Academic Press, New York

Nash, H. A., and Robertson, C. A. (1981) J. Biol. Chem. 256, 1069–1072

MacWilliams, M. P., Gumport, R. I., and Gardner, J. F. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 6104–6109

Hoess, R., Abremski, K., Irwin, S., Kendall, M., and Mack, A. (1990) J. Mol. Biol. 226, 126–131

Esposito, D., and Scocca, J. J. (1997) Nucleic Acids Res. 25, 3605–3614

Kwon, H. J., Tirumalai, R. S., Kwon, H. J., Ellenberger, T., and Landy, A. (1998) Nucleic Acids Res. 26, 391–406

Argos, W., Landy, A., Abremski, K., Egan, J. B., Haggard-Ljungquist, E., Hoess, R. H., Kahn, M. L., Kalionis, W., Narayana, S. V. L., Pierson, L. S. I., Sternberg, N., and Leong, J. M. (1986) EMBO J. 5, 433–440

Blakely, G. W., and Sherratt, D. J. (1996) Mol. Microbiol. 20, 234–237

Chen, J.-W., Lee, J., and Jayaram, M. (1992) Cell 69, 647–658

Pan, G., Luetke, K., and Sadowski, P. D. (1993) Mol. Cell. Biol. 13, 3167–3175

Pargellis, C. A., Nunes-Duby, S. E., Meitoso de Vargas, L., and Landy, A. (1988) J. Biol. Chem. 263, 7678–7685

Lange-Gustafson, B. J., and Nash, H. A. (1984) J. Biol. Chem. 259, 12724–12732

Kleemann, P. (1986) EMBO J. 5, 1389–1393

Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R., and Davies, D. R. (1994) Science 266, 1981–1986

Subramanya, H. S., Ariszewska, L. K., Baker, R. A., Bird, L. E., Sherratt, D. J., and Wigley, D. B. (1997) EMBO J. 16, 5178–5187

Gor, F., Gepaol, D. N., and Van Dyme, G. D. (1997) Nature 390, 40–46

Tirumalai, R. S., Kwon, H., Cardente, E., Ellenberger, T., and Landy, A. (1998) J. Mol. Biol. 279, 513–527

Dorgai, L., Yapil, E., and Weisberg, R. A. (1998) J. Mol. Biol. 253, 178–188

MacWilliams, M. P., Gumport, R. I., and Gardner, J. F. (1996) Genetics 145, 1069–1079

Hoess, R., Abremski, K., Irwin, S., Kendall, M., and Mack, A. (1990) J. Mol. Biol. 214, 517–518

Panigrahi, G. B., and Sadowski, P. D. (1994) J. Biol. Chem. 269, 10940–10945

Nash, H. A., and Robertson, C. A. (1981) J. Biol. Chem. 256, 9246–9253

Borrowie, J. A., Zhang, L., Sasse-Dwight, S., and Gralla, J. D. (1987) J. Mol. Biol. 196, 101–111

Abremski, K., and Gottesman, S. (1982) J. Biol. Chem. 257, 9658–9662

Yin, S., Bushman, W., and Landy, A. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 1040–1044

Bradford, M. M. (1976) Anal. Biochem. 72, 248–254

Edman, P., and Begg, G. (1967) Eur. J. Biochem. 1, 80–91

Sarkar, D., Raman-Livaja, M., and Landy, A. (2000) EMBO J. 20, 1203–1212

Nunes-Duby, S. E., Matsumoto, L., and Landy, A. (1987) Cell 50, 779–788

Ross, W., and Landy, A. (1985) Cell 33, 261–272

Williams, K. R., and Konigsberg, W. H. (1991) Methods Enzymol. 208, 516–539

Allen, G. (1989) in Sequencing of Proteins and Peptides (Burdon, R. H., and van Kippenberg, P. H., eds), pp. 1–414, Elsevier, Oxford

Favre, A. (1990) in Bioorganic Photochemistry (Morrison, H., ed), pp. 379–425, John Wiley & Sons, Inc., New York

Meisenheimer, K. M., and Koch, T. H. (2000) Crit. Rev. Biochem. Mol. Biol. 35, 101–140

Frischauf, A. M., and Scheit, K. H. (1973) Biochem. Biophys. Res. Commun. 53, 1227–1233

Nikiforov, T. T., and Connolly, B. A. (1999) Nucleic Acids Res. 20, 1299–1314

McLaurathy, M. J., Boocock, M. R., and Stark, W. M. (1996) J. Mol. Biol. 260, 299–303

Fourrey, J. L., and Jolin, P. (1973) Tetrahedron Lett. 34, 3225–3227

Fourrey, J. L. (1976) Tetrahedron Lett. 4, 297–300

Pleiss, M., Ochiai, H., and Cerutti, P. A. (1996) Biochem. Biophys. Res. Commun. 34, 70–76

Pleiss, M. G., and Cerutti, P. A. (1971) Biochemistry 10, 3093–3099

Willis, M. C., Hicke, B. J., Uhlenbeck, O. C., Cech, T. R., and Koch, T. H. (1993) Science 262, 1255–1257

Cheng, Q., Swalla, B. M., Beck, M., Alcaraz, R., Jr., Gumport, R. I., and Gardner, J. F. (2000) Mol. Microbiol. 36, 424–436

Chen, Y., Narendra, U., Iype, L. E., Cox, M. M., and Rice, P. A. (2000) Mol. Cell 6, 888–897

Erskine, S. G., and Halford, S. E. (1998) J. Mol. Biol. 275, 759–772

Azaro, M. A., and Landy, A. (2001) in Mobile DNA II (Craig, N. L., Craigie, R., Gellert, M., and Lambowitz, A., eds), pp. 118–148, American Society for Microbiology, Washington, D. C.

Dorgai, L., Sloan, S., and Weisberg, R. A. (1998) J. Mol. Biol. 277, 1059–1070