doi:10.1016/S0022-2836(03)00469-8 J. Mol. Biol. (2003) 329, 423–439

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

# Dynamics and DNA Substrate Recognition by the Catalytic Domain of Lambda Integrase

**Srisunder Subramaniam[1], Arun K. Tewari[2], Simone E. Nunes-Duby[3] and Mark P. Foster[1,2]***

[1]*Biophysics Program, The Ohio State University, Columbus OH 43210, USA*

[2]*Department of Biochemistry The Ohio State University, 484 West 12th Ave., Columbus, OH 43210, USA*

[3]*Division of Biology and Medicine, Brown University Providence, RI 02912, USA*

**Corresponding author*

Bacteriophage lambda integrase (λ-Int) is the prototypical member of a large family of enzymes that catalyze site-specific DNA recombination *via* the formation of a Holliday junction intermediate. DNA strand cleavage by λ-Int is mediated by nucleophilic attack on the scissile phosphate by a conserved tyrosine residue, forming an intermediate with the enzyme covalently attached to the 3′-end of the cleaved strand *via* a phosphotyrosine linkage. The crystal structure of the catalytic domain of λ-Int (C170) obtained in the absence of DNA revealed the tyrosine nucleophile at the protein's C terminus to be located on a β-hairpin far from the other conserved catalytic residues and adjacent to a disordered loop. This observation suggested that a conformational change in the C terminus of the protein was required to generate the active site in *cis*, or alternatively, that the active site could be completed in *trans* by donation of the tyrosine nucleophile from a neighboring molecule in the recombining synapse. We used NMR spectroscopy together with limited proteolysis to examine the dynamics of the λ-Int catalytic domain in the presence and absence of DNA half-site substrates with the goal of characterizing the expected conformational change. Although the C terminus is indeed flexible in the absence of DNA, we find that conformational changes in the tyrosine-containing β-hairpin are not coupled to DNA binding. To gain structural insights into C170/DNA complexes, we took advantage of mechanistic conservation with Cre and Flp recombinases to model C170 in half-site and tetrameric Holliday junction complexes. Although the models do not reveal the nature of the conformational change required for *cis* cleavage, they are consistent with much of the available experimental data and provide new insights into the how *trans* complementation could be accommodated.



*Keywords:* lambda integrase; covalent protein/DNA complex; proteolysis; proteolysis; NMR

## Introduction

Bacteriophage λ-integrase (λ-Int) is a multi-domain enzyme that, with the assistance of host- and phage-encoded accessory protein factors (Xis, FIS and IHF), is responsible for catalyzing site-specific insertion and excision of the bacteriophage viral DNA into and out of the *E. coli* host genome, enabling the lysogenic and lytic phases of the phage life cycle. The enzyme catalyzes site specific recombination *via* a highly orchestrated series of events involving assembly of a tetrameric structure, coordinated cleavage of two opposing DNA strands to produce a covalent protein–DNA complex, followed by strand transfer and formation of a Holliday junction intermediate; a central isomerization step is thought to change the geometry of the Holliday junction intermediate and determine the direction by which the synapse is resolved (Figure 1).[1–12] Although much is known about the overall mechanism of the reactions catalyzed by

Present address: A. K. Tewari, Physiology & Cell Biology, 304 Hamilton Hall, 1645 Neil Ave., Columbus, OH 43210, USA.

Abbreviations used: λ-Int, bacteriophage lambda integrase; C170, λ-Int catalytic-domain comprising residues 170–356.

E-mail address of the corresponding author: foster.281@osu.edu

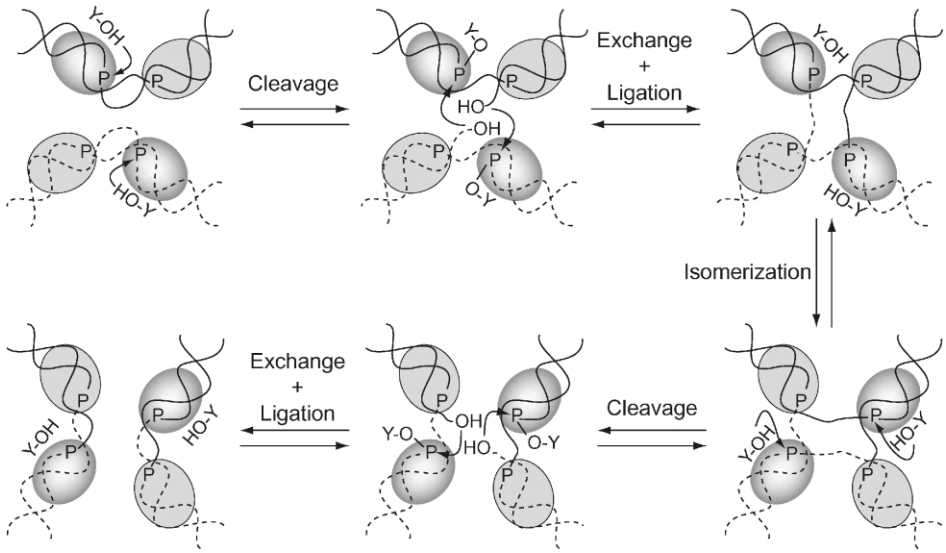



**Figure 1**. The mechanism of site-specific DNA recombination mediated by bacteriophage λ integrase (λ-Int).[9] The enzyme recognizes a pair of semi-symmetric inverted sequence elements at the site of recombination, one from the phage DNA and one from the host genome (dashed, solid lines). The activated pair of λ-Int protomers that mediates DNA strand cleavages is highlighted. During strand cleavage a conserved tyrosine residue (Y-OH) performs nucleophilic attack on the scissile phosphate (P) to yield a covalent protein–DNA complex intermediate with the tyrosine covalently attached to the 3′-phosphate of the cleaved strand (Y-O–P). This is followed by the exchange and ligation of the cleaved strands to generate a Holliday junction intermediate. In this mechanism, protein–protein interactions mediate a critical isomerization step that changes which pair of protomers is activated for DNA cleavage. Once isomerized, a second set of strand cleavages, exchanges and ligations yield the fully integrated (or excised) product.

λ-Int and related site-specific recombinases, the nature of the conformational changes in the proteins and protein–DNA complexes that facilitate the various steps in the pathway remain poorly understood.

λ-Int possesses three distinct structural domains: an N-terminal domain (residues 1–64) responsible for binding to distal *arm* sites on the phage DNA,[13–15] a central domain (CB; 65–169) required for high-affinity binding to core-type sites,[16] and the C-terminal catalytic domain (170–356).[11,17] Although the catalytic domain on its own cannot carry out recombination, it is competent in specifically recognizing and cleaving its cognate core DNA substrates, either as full- or half-sites.[17]

The enzyme active site is composed of a conserved tetrad consisting of Arg, His-X-X-Arg and Tyr (Figure 2), with the first three residues serving to activate the scissile phosphate for nucleophilic attack by the tyrosine hydroxyl leading to esterification of the 3′-phosphoryl group of the recombination site.[18–20] λ-Int and other members of the tyrosine recombinase family share mechanistic similarity with type I topoisomerases[21] in that nucleophilic attack by the conserved tyrosine residue on the scissile phosphodiester linkage yields a covalent 3′-phosphotyrosine intermediate and a free 5′-OH on the 3′ nucleotide.

The crystal structure of the catalytic domain of λ-Int (residues 170–356; C170)[11] revealed that while most of the residues constituting the conserved catalytic core are concentrated in the center of the molecule, the tyrosine nucleophile ($Tyr^{342}$, mutated to Phe in the crystallographic study, C170-Y342F) was located on a β-hairpin far (>18 Å) from the other residues in the catalytic tetrad ($Arg^{212}$, $His^{308}$ and $Arg^{311}$) (Figure 2). Furthermore, no density was observed for an eight-residue loop ($Lys^{334}$ to $Gln^{341}$) adjacent to the catalytic tyrosine that coincides with a region of the protein previously shown to be proteolytically sensitive.[17] The significance of the observed arrangement of the catalytic residues remained unclear, although the flexibility implied by the crystallographic disorder and proteolytic sensitivity led to the proposal that the crystal structure might represent an *open*, inactive, conformation, and that λ-Int could undergo a binding-coupled conformational change in an induced fit manner to generate a *closed*, active conformation.[11,17,22] In addition, the finding was also attractive in that it provided a plausible explanation for the apparent mechanistic duality[23,24] in which the active site tyrosine of one λ-Int molecule could either pair in *trans* with the active site residues of another molecule to catalyze DNA strand cleavage,[25] or in *cis* with its own active site residues.[22,26,27]

Subsequently, the crystal structures of Int-family members, bacteriophage P1 Cre and yeast FLP recombinases, in covalent and non-covalent complexes with their respective DNA substrates, provided exquisite insights into the structures of synaptic complexes and Holliday junction recombination intermediates.[28–32] In the structure of Cre recombinase bound to a symmetrized substrate, the active site is clearly composed of residues

from one protomer, and is thus consistent only with *cis* cleavage.[28,29] In contrast, the crystal structure of yeast FLP recombinase in a tetrameric structure bound to DNA revealed *trans* assembly of the active site,[31,32] consistent with complementation experiments.[33–35]

We used NMR and biochemical methods to study the dynamics of the catalytic domain of λ-Int (C170) in solution, in the absence and presence of DNA substrates with the objective of exploring the nature of the conformational changes in C170 that accompany DNA binding and catalysis. In addition, we have taken advantage of the strict conservation of the active site residues in the tyrosine recombinase family[20] to construct an atomic three-dimensional model of the λ-Int catalytic domain in monomeric and tetrameric complexes with DNA by docking the crystallographically determined structure of C170 onto the structures of Cre and Flp recombinases. These models, constructed without bias towards the question of *cis*/*trans* cleavage, are consistent with our spectroscopic and proteolytic data and may provide some unique and novel insights into λ-Int function.

## Results and Discussion

The domain boundaries in λ-Int were identified previously.[13,16,17] C170-Int (C170), the recombinant construct containing the catalytic domain of λ-Int used for these studies,[16,17] consists of 187 residues: 170–356 (Figure 2(a)); the first ordered residue in the crystal structure of C170-Y342F was Arg177.[11] In spite of low sequence identity, the secondary structures of the catalytic domains of λ-Int and Cre (as well as other family members) are highly conserved (Figure 2(b)). One of the most notable differences between the secondary structures of the two proteins is the absence in λ-Int of the helical segments at the C terminus of Cre that contain the tyrosine nucleophile (αM) and mediate protein–protein interactions in the recombining synapse (αN).[20,29]

The core recombination DNA sequences recognized by λ-integrase on the phage and host bacterial genomes, termed *attP* and *attB*, respectively are shown in Figure 3(a); upon integrative recombination the products are termed *attL* and *attR*, for the left and right segments of the prophage DNA. The two core recognition half-sites on the phage DNA are termed *C* and *C′* sites, and the corresponding sequences on the bacterial DNA are termed *B* and *B′*.[5] In Figure 3 the sequences are numbered such that the nucleotide at the center of the 7-bp overlap region between the cleavage sites is "0", nucleotide numbers decrease to the left and increase to the right (5′ to 3′). With this notation, the scissile phosphate is located between bases −4 and −3 on each of the cleaved strands, and the core recognition sequences span nucleotides −10 to −2 on each cleaved strand; the nucleotides on the complementary strand are 2 through 10.

Productive interaction between λ-Int and a pair of core sequences involves assembly of a (large) tetrameric Holliday junction intermediate comprising four protein molecules and their respective binding sites on two conjoined DNA oligonucleotides. To reduce the size of the protein–DNA complex and yet study a catalytically-relevant interaction, we took advantage of the observation that C170 (like full-length λ-Int) can cleave "half-site" DNA substrates, to examine the interaction between C170 and substrates encompassing either the *B′* site or a hybrid between the *C* and *C′* sites.[18,36,37] The minimal λ-Int half-site substrate is a 13-mer (spanning nucleotides −13 to −1)[22] that generates a trapped intermediate because of the irreversible diffusion of the short −3 to −1 oligomer that results upon strand cleavage.[18,22,36] However, these suicide complexes are susceptible to irreversible hydrolysis[38] and are thus unstable on the time-scale required for NMR experiments. In order to characterize the protein in covalent and non-covalent complexes with half-site substrates, two thermodynamically stabilized DNA hairpins[39–42] were used (Figure 3(b)).

### Limited proteolysis of covalent and non-covalent C170-DNA complexes

Limited proteolysis experiments with ArgC and kallikrein previously demonstrated the accessibility of the carboxyl-terminal region of C170.[17] We performed similar experiments with trypsin (Figure 4(a)), and from electrospray mass spectrometric analysis of partially digested C170 identified two primary cleavage sites. The mass of the full-length protein (after removal of the translationally encoded amino-terminal methionine) is 21,119 Da, as confirmed by mass analysis of the protein prior to trypsin cleavage. Limited trypsin proteolysis yielded two peptides, one with a mass of 20,848 Da, corresponding to residues 173–356 (20,848.80 Da predicted), and a second peptide with a mass of 19,206 Da, corresponding to residues 173–343 (19,207 Da predicted). This result identifies the primary trypsin cleavage sites as $Lys^{172}$ and $Arg^{343}$ (indicated in Figure 2 by "†") and confirms the accessibility of the protein in the vicinity of the tyrosine nucleophile ($Tyr^{342}$).[17,43] Furthermore, because the tyrosine nucleophile is retained in both large tryptic fragments (residues 173–356 and 173–343), limited proteolysis of the protein covalently bound to DNA should identify whether the C-terminal loop is accessible to trypsin in that state.[43]

If conformational changes in the protein from an "open" inactive state to a "closed" active state were coupled to DNA binding, thereby bringing the tyrosine nucleophile into the vicinity of the other catalytic residues and into position for attack on the DNA substrate, we might expect reduced proteolytic accessibility of the C-terminal hairpin (at

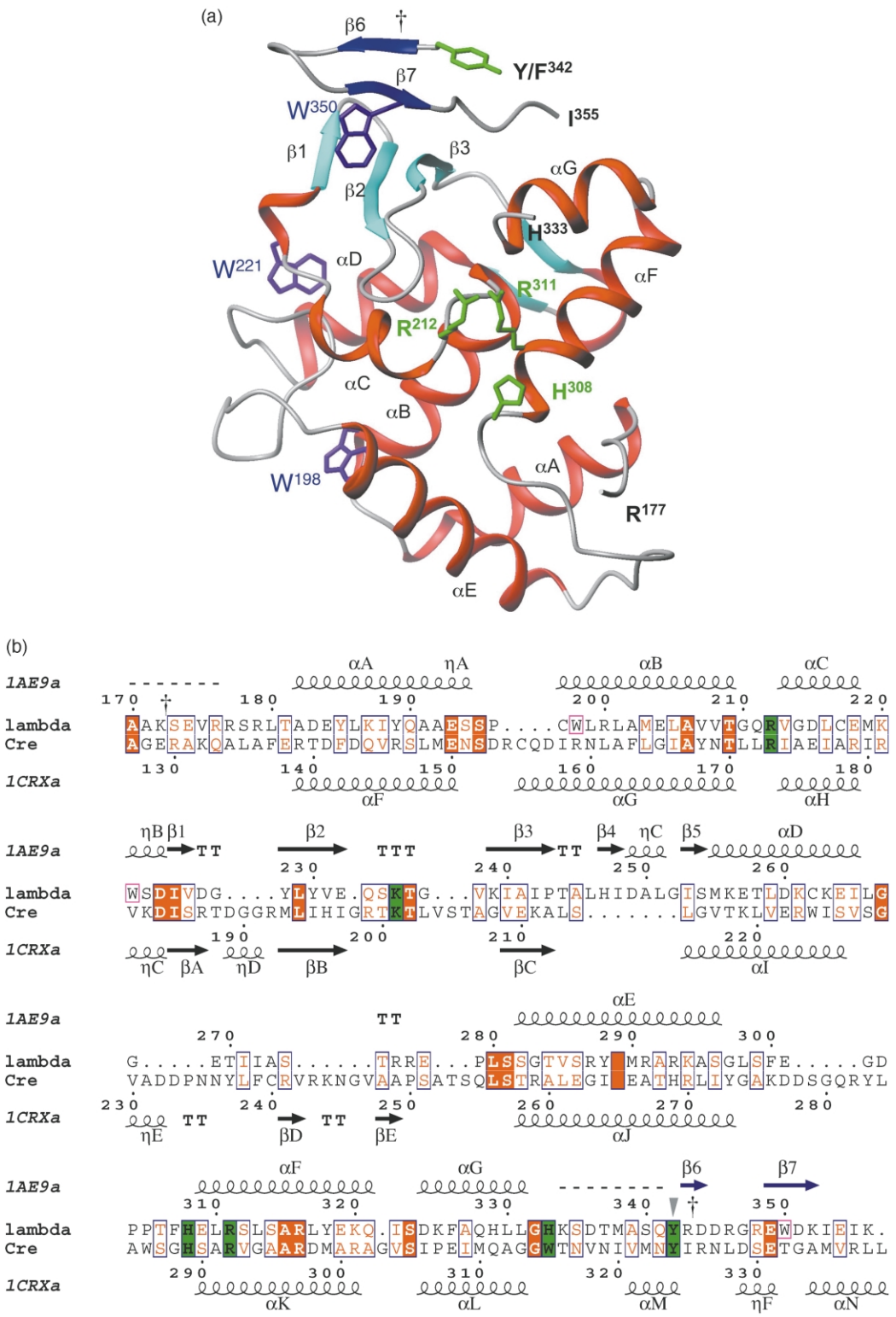



**Figure 2**. Structure of the λ-Int catalytic domain. (a) Ribbon diagram of the λ-integrase catalytic domain based on the crystal structure of C170-Y342F, a cleavage-deficient mutant comprising the C-terminal 187 residues of λ-Int (170–356).[11,17] The key conserved catalytic residues are green while the three tryptophan residues in C170 are magenta. The helices are red, strands cyan, and the hairpin containing the tyrosine nucleophile (a Phe in 1AE9) is dark blue. No electron density was observed for the loop spanning residues $His^{333}$ to $Tyr/Phe^{342}$. The site of trypsin cleavage on the C-terminal hairpin of free C170 ($Arg^{343}$) is indicated by "†".[11] (b) Sequence of the catalytic domain (C170) of λ-Int (P03700) aligned with the catalytic domain of Cre recombinase (P06956). Secondary structures indicated are as designated in the PDB entries 1EA9A and 1CRXA; $3_{10}$ helices are labeled η. Residues to which no density was assigned in the crystal structure of C170 (residues 170–176, 334–341, 356) are indicated by "--". The alignment shown has been described previously[20] and was generated according to sequence and structural homology; the figure was generated with ESPript (http://maple.bioc.

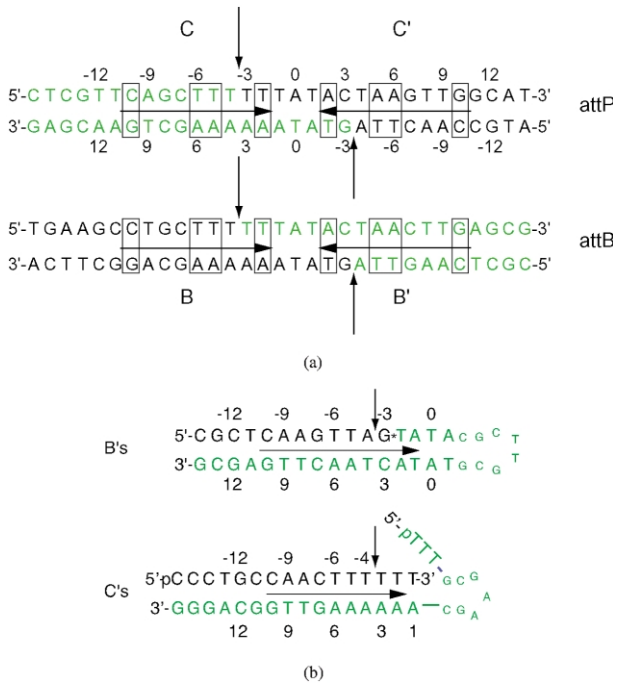



**Figure 3**. λ-Int core DNA substrates. (a) Core recognition sequence of phage (*attP*), and bacterial (*attB*) attachment sites. The cleavage sites are indicated by the vertical arrows. Recombination joins the green (*C*, *B*′) and black (*B*, *C*′) fragments to form the *attL* and *attR* products. The invariant nucleotides (boxed) are at positions −10, −6, −5 and −2 on each of the cleaved strands. (b) Half-site core recognition substrates used in this work. *B*′s is a suicide substrate that contains the *B*′ recognition sequence and has a nick between the −3 and −2 nucleotides (indicated by "∗"). *C*′s is a suicide substrate containing the *C*′/C consensus site designed such that upon cleavage by λ-Int, the TTT immediately downstream of the cleavage site will be lost to diffusion, and the 5′-pTTT from the complementary oligomer will anneal in its place. Because the bottom strand has been phosphorylated at the 5′-end, the 5′-OH that normally would displace the tyrosine residue in the phosphotyrosine intermediate is not present, trapping the covalent complex. The 5′-phosphate at the recombining site prevents 5′OH-mediated re-ligation and also increases the stability of the covalent intermediate by blocking access to the active site by water.[4,22,37,38]

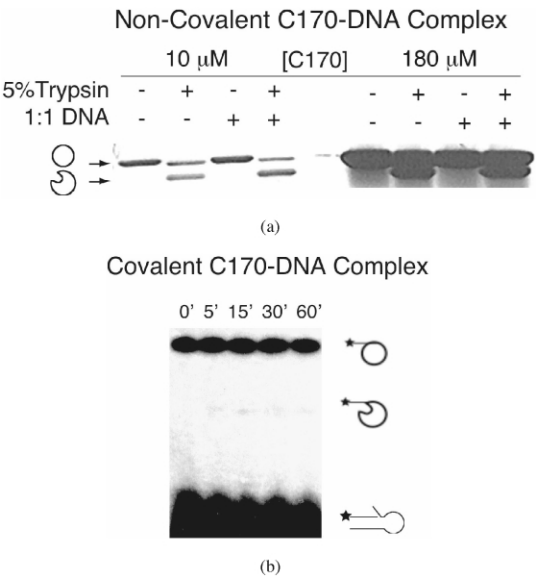



**Figure 4**. Limited trypsin proteolysis of C170. (a) Coomasie stained SDS-PAGE analysis of trypsin-treated C170 free and bound (non-covalently) to DNA. For lanes with DNA, prior to a 5-minute trypsin digestion, the protein (at 10 μM or 180 μM) was incubated at a 1:1 stoichiometry with the *C*′s substrate for 15 minutes, long enough for binding, but not long enough to generate measurable quantities of covalently trapped protein. The gel shows two bands after cleavage with trypsin; the upper band consists of both full-length C170 and the unresolved proteolytically-generated fragment comprising residues 173–356 (after loss of the three N-terminal residues), while the lower band corresponds to the 173–343 fragment generated by cleavage immediately following the tyrosine nucleophile in the C-terminal hairpin. (b) Analysis of the proteolytic sensitivity of C170 covalently trapped to the [5′-$^{32}$P]-labeled *C*′s substrate. The protein was incubated (in a 100-fold molar excess) with the radiolabeled *C*′s substrate for two days at room temperature prior to digestion for the indicated time with trypsin. Cleavage products were separated by 18% SDS-PAGE and visualized by autoradiography.

Arg[343]) upon addition of a DNA substrate. However, trypsin cleavage of non-covalent protein/DNA complexes (obtained after brief incubation of C170 with DNA substrate; see Materials and Methods) revealed no significant reduction in accessibility of the loop (Figure 4(a)). This result implies that either the expected conformational change in the C terminus of C170 is not strongly coupled to DNA binding, the conformational change is transient and not significantly populated, and/or the new conformation is equally accessible to the protease.

In contrast, when we probed the accessibility of the covalent C170–DNA complex (obtained after prolonged incubation with a large excess of protein; see Materials and Methods) we found it to be highly resistant to proteolysis (Figure 4(b)). Thus, the C terminus of C170, when in a covalent complex with substrate (i.e. resembling the active

columbia.edu/cgi-bin/pp/nph-ESPript_exe.cgi),[80] with identity indicated by blocked letters (red) and those with a Risler similarity score of >0.7 by boxes. The conserved catalytic residues are green, the tyrosine nucleophile is indicated with an inverted triangle, the invariant Arg, His-xx-Arg with "∗", the three tryptophan residues in C170 are boxed in magenta, and the sites of trypsin cleavage are indicated by "†".

conformation) is in a conformation inaccessible to the protease. This result clearly demonstrates that the covalent and non-covalent complexes differ in this respect, but does not provide detail into the structural and/or dynamic nature of this difference.

### NMR resonance assignments of C170

We sought to obtain a detailed description of the substrate-induced conformational change by NMR spectroscopy. Complete backbone resonance assignments of free C170 have been complicated by $C^{\alpha}$ resonance overlap, low solubility and aggregation. The latter two problems have contributed to low sensitivity in triple resonance experiments that involve multiple magnetization transfer steps (i.e. HNCACB, CBCA(CO)NH). Consequently, while work is ongoing to enable complete resonance assignments by stabilizing the protein and improving spectral quality, we have proceeded with the available partial assignments to gain important functional insights. Although interpretation of data with partial resonance assignments carries the risk that some resonances could be miss-assigned, particularly when there is chemical shift degeneracy, the risk is limited for backbone resonances when the assignments are based on through-bond connectivities to the preceding and following residue in the sequence; only backbone assignments with at least one such unambiguous correlation have been used in this analysis.

Analysis of data from 3D HNCA, HN(CO)CA, HNCACB and CBCA(CO)NH spectra, together with careful analysis of backbone chemical shifts,[44,45] allowed us to assign the backbone amide resonances of 14 fragments comprising 65 of the 187 residues in C170: 172–176, 191–195, 207–215, 228–229, 237–242, 251–253, 267–270, 273–274, 283–284, 295–303, 331–333, 336–341, 344–348 and 354–356 (as indicated in Figure 5(a)). Most of these assigned segments correspond to sequences with unique patterns of $C^{\alpha}$ shifts or reflect portions of the protein that are more mobile on average than the rest of the protein, resulting in slower transverse ($T_2$) relaxation and narrower resonance lines (below). Importantly, these assignments allow us to obtain critical insights into the dynamics and effect of DNA binding on the region of the protein comprising the crystallographically disordered loop between helix G and strand 6 (residues ~333–341) and the C-terminal hairpin containing the tyrosine nucleophile. In addition to the partial backbone resonance assignments, the side-chain indole $N^{\varepsilon 1}$ signals of the three tryptophan residues in C170 ($Trp^{198}$, $Trp^{221}$ and $Trp^{350}$; see Figure 2) are well resolved (Figures 5(a) and 6) and serve as site-specific probes of the dynamics and environment of the protein (see Materials and Methods for assignment of these resonances). Note that $Trp^{198}$ is at a position that in the Cre and Flp recombinase crystal structures mediates a phosphate contact to DNA,[28–31] while $Trp^{221}$ is buried in the protein core.

### C170 dynamics from $^{15}N$ relaxation

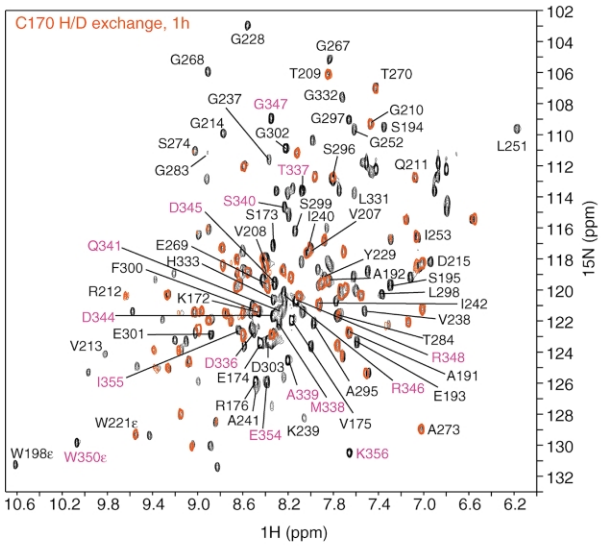
NMR data indicate that C170 is highly flexible in the absence of DNA, and exhibits a wide distribution of motional time-scales. For instance, very few of the assigned resonances are significantly protected from hydrogen/deuterium (H/D) exchange (Figure 5(a)), including most notably, those in the C terminus of C170 (residues 331–356). Since many of the assigned regions are in loops, their lack of protection is expected, but the fact that the amides in the C-terminal hairpin bearing the tyrosine nucleophile also exchange rapidly indicates that this structural element is at best transient in solution.

In addition to the time-scales sampled by H/D exchange, the time-scales probed by $^{15}N$ relaxation also reveal substantial dynamics in the free protein. The 10% trimmed mean values for the $^{15}N$ $T_1$, $T_2$ and NOE were 1.4 ± 0.3 seconds, 49 ± 9 ms and 0.82 ± 0.16, respectively (Figure 5(b)). The broad distribution of $^{15}N$ $T_1/T_2$ ratios indicate the protein has substantial regions that are highly mobile on the ms–μs time-scale, [46,47] while the heteronuclear $\{^1H\}-^{15}N$ NOE data indicate that some of this flexibility is manifested on the ns–ps time-scale. Although the protein has some tendency to aggregate and or dimerize in solution,[48] the overall rotational correlation time, $\tau_e$, calculated from the $T_1/T_2$ ratio (~30) was ~12.5 ns, close to that expected for a C170 monomer.[47,49]

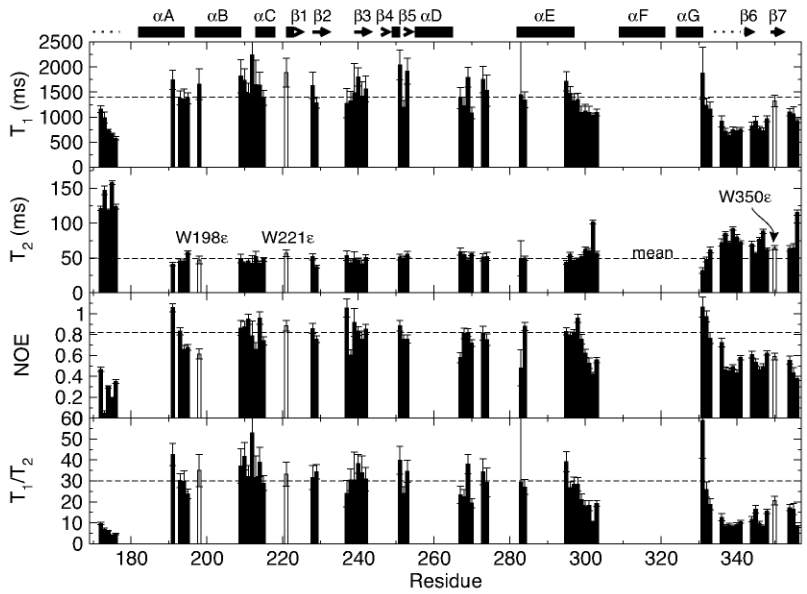
From the amide $^{15}N$ relaxation rates (Figure 5(b)) of the assigned residues, highly flexible regions of the protein can be readily identified. In particular, it is clear that the extreme amino-terminal segment (residues 172–176), the loop between helices E and F (residues 295–303), and residues at the extreme C terminus of the protein where the catalytic tyrosine is located (residues 336–356) are highly mobile, as indicated by small $\{^1H\}-^{15}N$ NOE values and $T_1/T_2$ ratios. This is in agreement with the observed disorder or higher temperature factors of these regions in the crystal structure of C170-Y342F[11] and with their proteolytic sensitivity, as previously observed[17,43] and further demonstrated here. In addition, $Trp^{350}$, which serves as an additional indicator of the dynamics of the C-terminal hairpin has a shorter $T_1$, longer $T_2$ and smaller $\{^1H\}-^{15}N$ NOE compared to the other two tryptophan residues. Thus, consistent with its proteolytic sensitivity and crystallographic disorder, the $^{15}N$ NMR relaxation data indicate the C-terminal region of the protein containing the tyrosine nucleophile is flexible in the absence of DNA.

### NMR studies of the C170-DNA non-covalent complex

To identify residues in C170 involved in DNA binding,[50] we recorded an $^{15}N$ HSQC spectrum on




(a)




(b)

**Figure 5**. Dynamic heterogeneity of C170 on a range of time-scales. (a) Overlay of the HSQC and one-hour time point for H/D exchange of C170; partial assignments are as indicated with labels for crosspeaks from the flexible C-terminus in magenta. (b) $^{15}$N relaxation data for assigned $^{1}$H–$^{15}$N resonances in C170. The filled bars indicate the data for backbone amides, while the three side-chain tryptophan indole $H^{\varepsilon 1}$–$N^{\varepsilon 1}$ signals ($W^{198}$, $W^{221}$ and $W^{350}$) are indicated by the open bars. The mean values are indicated with the horizontal dashed lines. Secondary structural elements are indicated as in Figure 1(b).

a 1:1 non-covalent complex between C170 the *B*'s half-site DNA substrate. As can be seen from an overlay of the spectra from free C170 and the C170/*B*'s complex (Figure 6), significant shift perturbations and/or broadening were observed for the backbone amides of Lys$^{172}$-Arg$^{176}$, Arg$^{212}$-Gly$^{214}$, Gly$^{237}$-Val$^{238}$, Gly$^{267}$-Thr$^{270}$, Ser$^{296}$, Ser$^{299}$-Gly$^{302}$ Leu$^{331}$-Gly$^{332}$ and the side-chain of Trp$^{198}$. Although the available assignments do not permit constructing a complete interaction map, the shift perturbations are consistent with the protein binding to DNA in a manner analogous to that of related family members, Cre and Flp (below).

What is most remarkable in the shift perturbation data is that the resonances in the C-terminal hairpin are completely unaffected by binding to a half-site substrate. Compare for instance (Figure 6) the non-perturbed resonances from the C-terminal hairpin (Thr$^{337}$, Ala$^{339}$, Ser$^{340}$, Glu$^{354}$, Lys$^{356}$ and the Trp$^{350}$ side-chain indole) to the perturbed resonances of residues in the αD-αE loop (Gly$^{267}$, Gly$^{268}$, Thr$^{270}$), the end of helix G (Leu$^{331}$, Gly$^{332}$), the β2-β3 loop (Gly$^{237}$, Val$^{238}$), the αB-αC loop (Val$^{213}$, Gly$^{214}$, Arg$^{212}$) and the side-chain indole of Trp$^{198}$.

From the C170-Y342F crystal structure (Figure 2(a))[11] it is evident that in order for *cis* cleavage to occur, a substantial conformational change (from *open* to *closed*) would be required to bring Tyr$^{342}$ into the vicinity of the other catalytic residues. Regardless of whether the C terminus is rapidly sampling the *open* and *closed* conformations or whether the new conformation is only induced

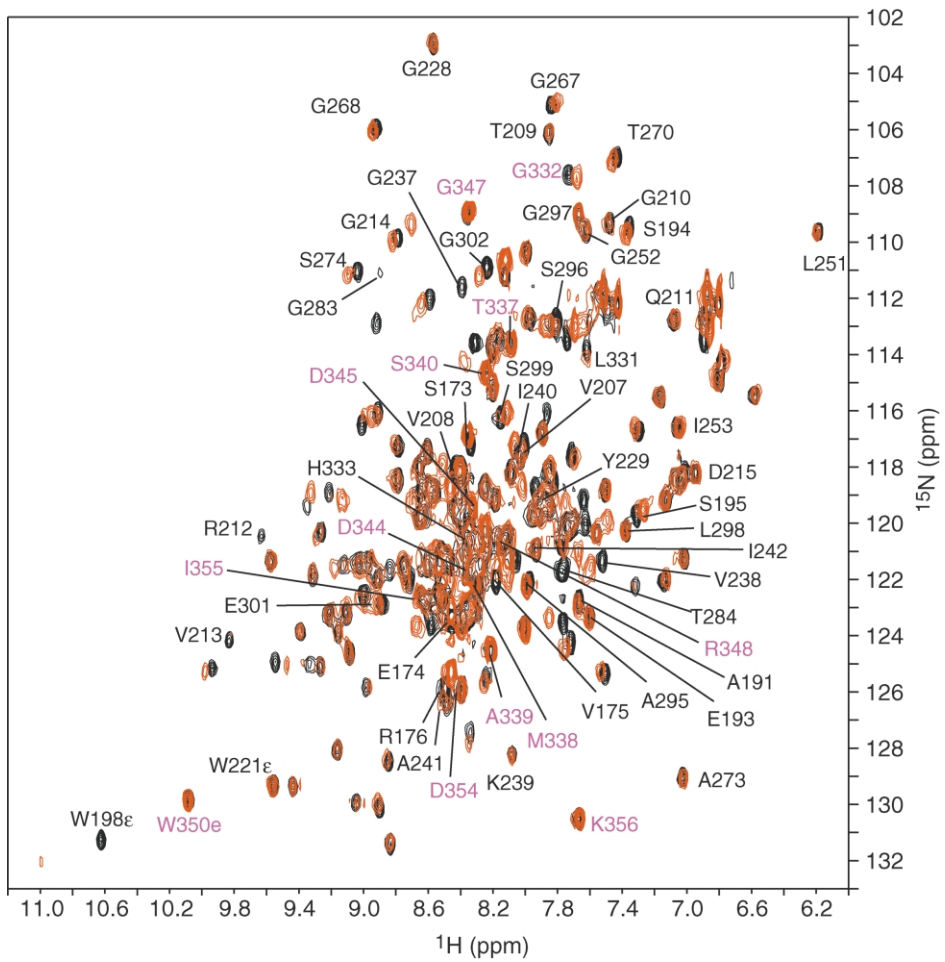



**Figure 6**. Overlay of the $^{15}$N HSQC spectra (~200 μM, 800 MHz 25 °C) of C170 free (black) and bound to the *B*′s DNA half-site (red). Partial resonance assignments of free C170 are as indicated.

in response (i.e. coupled) to DNA binding, such a conformational change would be expected to result in a dramatically different spectral signature for the residues in the C-terminal segment of the protein, including that of the well-resolved Trp$^{350}$ side-chain indole. Since no measurable perturbation was seen, it seems that either (1) the C-terminal hairpin is not involved in binding DNA half-sites, or (2) that the required *closed* conformation is present at a level below that which could be detected (less than ~5% of the population). In either case, the experimental data alone do not allow us to understand how this *catalytically competent* domain can cleave substrates (Figure 4(b)).[16,17]

## Docking model of the C170-DNA half-site complex

We took advantage of the structural and mechanistic conservation in the catalytic domains of λ-Int and Cre (Figure 2), to develop a three-dimensional docking model that could be used to interpret the available biochemical and spectroscopic data for the C170/half-site complex. This model was generated by a rigid-body superposition of the C$^{\alpha}$ atoms of Arg$^{212}$, His$^{308}$ and Arg$^{311}$ in λ-Int onto those of the corresponding residues in Cre (Arg$^{173}$, His$^{289}$ and Arg$^{292}$); the RMSD for the fit was 0.6 Å. The same superposition onto the C$^{\alpha}$ atom coordinates of Arg$^{190}$, His$^{305}$ and Arg$^{308}$ in the Flp/*fre* complex overlaid with an RMSD of 0.4 Å (not shown). Despite the fact that the coordinates of these essential catalytic residues were the only restraints in docking C170 onto DNA, the resulting model (Figure 7) is appealing in that it provides unique insights into data available in the literature and reported here.

The docking model exhibits very good overall architectural agreement with the Cre/*loxP* catalytic core structure, positioning helix E in the major groove for sequence recognition, [51–53] and exhibiting coincidence of all helical segments with those of Cre (Figure 7(a)). In addition to the excellent overall agreement, the model is also attractive at the residue level. For instance, although C170 was crystallized in the absence of DNA, the side-chains of the catalytic Arg, His-xx-Arg triad are oriented in a very similar manner to that seen in Cre where they are positioned for activation of the scissile

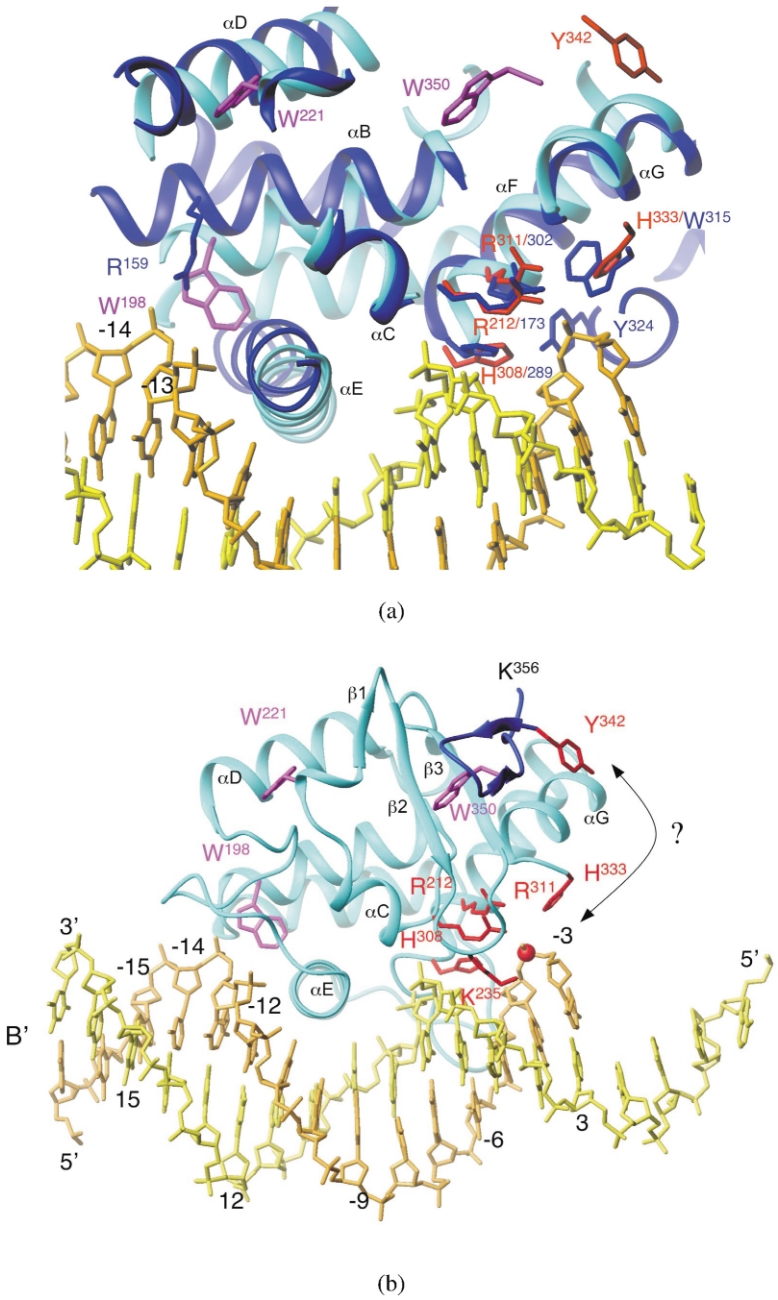



**Figure 7**. Modeling the C170/DNA half-site complex. (a) Superposition of C170 on the coordinates of the Cre/*loxP* complex.[28] The helices of C170 are cyan and the corresponding structural elements in Cre are dark blue; the tryptophan side-chains are magenta and the catalytic residues in C170 (including the tyrosine nucleophile, Tyr342, a Phe in the crystal structure, 1AE9[11]) are shown in red. (b) Model of C170 on a half-site B′ (*attB*) DNA substrate. The C170 core is represented as a cyan ribbon and the hairpin containing both the catalytic tyrosine and Trp$^{350}$ is dark blue. The DNA half-site is shown with the non-cleaved strand yellow and the cleaved strand orange; the scissile (−3) phosphate is indicated with a red sphere. The arrow indicates the conformational change required for formation of an intact active site in *cis*.

phosphate. This is also the case for two other important catalytic residues,[29] Lys$^{235}$ and His$^{333}$ (Lys$^{201}$ and Trp$^{315}$ in Cre; Figure 2(b)), even though these residues are in loops in C170 and were not used as constraints in docking. In contrast, the position of the tyrosine nucleophile in Cre (Tyr$^{324}$) can be seen poised for nucleophilic attack, whereas that of λ-Int (Tyr$^{342}$, a Phe in the crystal structure) is far from the scissile phosphate.

The position of the indole side-chain of the tryptophan residues was of particular interest in light of the observation that only one of the H$^{\varepsilon 1}$ resonances was perturbed by DNA binding (Figure 6). The side-chain of Trp$^{198}$, located on helix B, is perfectly positioned to contact the DNA phosphate backbone between nucleotides −13 and −14. In Cre, this position is occupied by Arg$^{159}$, which is indeed involved in a phosphate contact (Figure 7(a)); Tyr$^{176}$ makes the analogous contact in the Flp/*fre* complex. Although sequence identity at this position is low, in most members of the Int family, it is occupied by residues capable of mediating a phosphate contact *via* a hydrogen bond.[20] The other two tryptophan residues, including Trp350 on the C-terminal hairpin, are located far from the DNA interface in this model.

The DNA-induced chemical shift perturbations are also consistent with the half-site model (Figure 7(b)). Shift perturbations are observed for

residues at the extreme amino terminus (residues 171–176) and the flexible loop (residues 295–303) between recognition helix E and helix F, which bears the catalytic His$^{308}$-X-X-Arg$^{311}$ (located behind helix G in Figure 7(b)). These two regions are near each other and in the docking model are close to the DNA minor groove in the vicinity of nucleotide −6 on the cleaved strand (behind the DNA in Figure 7(b)). Spectral changes are observed for residues in the loop between helices B and C (containing catalytic Arg$^{212}$), residues 237–238 in the loop between strands 2 and 3 (near the catalytically important Lys$^{235}$), and residues 331–333 at the end of helix G (which includes the catalytic residue, His$^{333}$); in the model, these residues all cluster near the scissile phosphate, −3 to −4. Similarly, shift perturbations are observed for residues (267–270) in the loop between helices D and E, which in the model is positioned near the DNA phosphate backbone across the major grove 5′ to the cleavage site, making contacts to phosphates in the vicinity of −14 to −12. As discussed above, no shift perturbations are observed for the C-terminal hairpin (residues 337–356), which is far from the DNA in the half-site model.

In addition to agreement with the spectroscopic data, some additional features of the modeled half-site complex are worth noting. The model suggests that when presented with a half-site DNA substrate, C170 binds the DNA primarily *via* major groove contacts mediated by helix αE to bases in the −8 to −12 region as well as phosphate and/or minor groove contacts in two regions, corresponding to nucleotides −4 to −5 and −12 to −14. This is consistent with the observed methylation protection of *attP* and *attB* sites by full-length λ-Int of guanine N7 positions in the major groove from nucleotides −7 to −12, of adenine N3 in the minor groove at positions −5 to −6.[54] Also, the phosphate contacts predicted from the C170 model are in agreement with those seen in the catalytic domains of Cre and Flp recombinases.[55,56] Finally, the half-site model is consistent with much available mutagenesis data.[20,22,27] For instance, mutation of Thr$^{270}$ and Ser$^{274}$, both of which are in the αD-αE loop that in our model contacts the phosphate backbone in the range of −12 to −15, inactivate the enzyme, and mutation in Cre of the residue at the analogous position to Trp$^{198}$ in λ-Int, results in diminished core-binding.

The model of C170 non-covalently bound to the half-site DNA substrate is consistent with proteolysis, chemical shift and dynamics data. However, it provides no insight into the formation of an "active" protein conformation that can become covalently attached to the DNA substrate. As demonstrated previously[16,17] and illustrated in Figure 4(b),[16,17] C170 is catalytically competent and forms covalent complexes with suicide substrates. In these complexes, the accessibility of the C-terminal hairpin to trypsin is severely restricted. Therefore, a critical part of the picture is missing in the NMR data of non-covalently bound protein and model of C170 on a DNA half-site.

## Tetrameric model of the C170–DNA recombination intermediate

To gain further insight into a covalent complex, we examined the crystal structures of Cre and Flp recombinases trapped in tetrameric complexes with suicide-substrates (Figure 8(a)).[28–32,57] Each of these complexes corresponds to a recombination intermediate in which the DNA has assembled into a Holliday junction structure. The Cre and Flp complexes are quite similar, with the most notable difference being that the active site in the Cre complex is constituted by residues from within a single protomer (*cis*) while the tyrosine nucleophile in the Flp active site is provided by a neighboring subunit (*trans*).

We used the same docking strategy, as described above for a C170/DNA half-site model, to generate a tetrameric model of C170 on a Holliday junction intermediate by superposing the C$^{\alpha}$ carbons of the invariant Arg, His-xx-Arg of C170 onto each molecule of Cre in the tetrameric complex. The resulting model of C170 on a Holliday junction substrate is shown in Figure 8(a), middle, and a close-up in Figure 8(b). Although the spacing in the wild-type λ-Int core sequence is seven base-pairs, the models constructed using either the 6 bp of Cre (shown) or 8 bp of Flp (not shown), exhibit very similar geometry and especially, potential inter-subunit contacts.

Since in the crystal structure of C170 in the absence of DNA the tyrosine nucleophile (a Phe in C170-Y342F) is far from the active site, we were particularly interested to see where the residue would be located in the tetrameric model. As can be seen in Figure 8, the C-terminal hairpin of C170, containing Phe/Tyr$^{342}$, is located in the inter-subunit interface where it might contact the adjacent subunit *via* residues in the flexible αE-αF loop and the linker preceding the structured catalytic domain. With the C-terminus as positioned, potential protein–protein contacts in the tetrameric model appear in a clockwise manner, as seen for Flp (clockwise: A′–B′–A–B), and contrary to that in Cre (counterclockwise: A′–B–A–B′).†

† After submission of this manuscript, a report appeared in the literature describing a model of a C170/DNA complex built by threading the λ-Int sequence into the Cre structure.[58] Because that model was based on the assumption that the C terminus of λ-Int rearranges to adopt the same structure as does Cre, the geometry of the inter-protomer contacts in the threading model differs from that obtained by the more conservative strategy of docking the experimentally determined structure onto DNA, as presented here. Nevertheless, our findings are also consistent with the experimental data and overall conclusions presented in that paper that indicate a role for the C terminus in mediating inter-protomer interactions and regulating recombination.

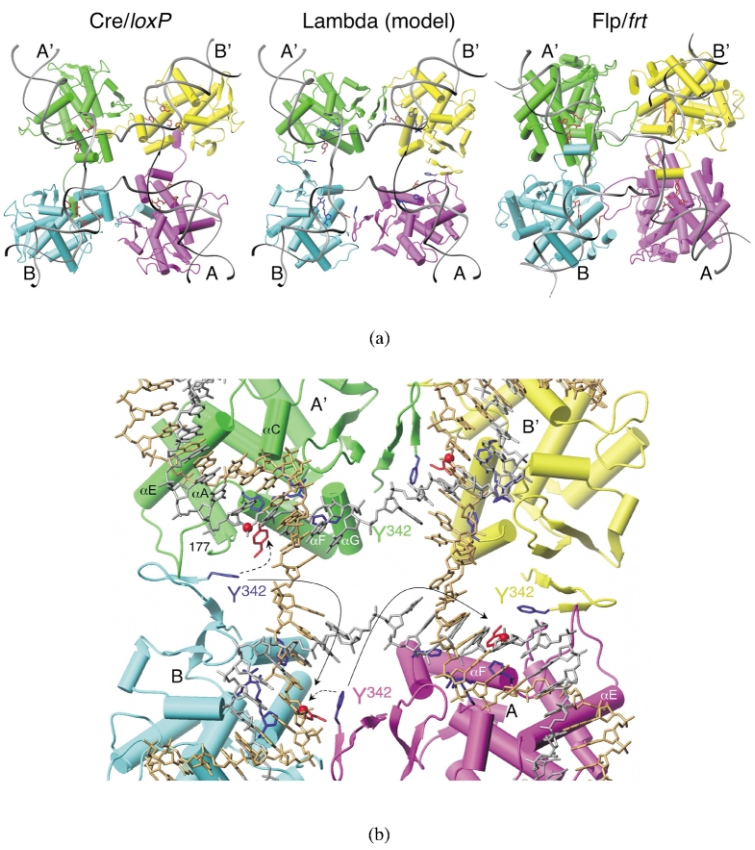



**Figure 8**. Tetrameric model of C170 in a recombining synapse. (a) Ribbon diagrams of the tetrameric structures of Cre (left) and Flp (right) recombinases in complex with symmetrized Holliday junction substrates.[28,31] Each protomer in the complex is color-coded, the protein helical segments are shown schematically with cylinders, and side-chains of the catalytic residues are shown as sticks; the DNA backbone is shown as a grey ribbon. For clarity and comparison to C170, only the catalytic domains of Cre and Flp are shown. The model of C170 in a recombining synapse is shown (center) in the identical orientation as Cre. (b) Close-up of the synapse in the tetrameric C170/Holliday junction model. The λ-Int catalytic residues are shown in blue, and the position of the tyrosine nucleophile in Cre is indicated in red. The DNA is indicated in grey and orange sticks, with each of the four the scissile phosphate groups highlighted in red. For molecules A and B (magenta, blue), the solid line indicates the conformational rearrangement required to generate *cis* cleavage, while the dashed lines indicate the conformational changes in that would enable *trans* cleavage.

An inspection of the interface between A′–B′ or A–B protomers shows that very little change in the conformation of the C terminus of C170 would be required to generate the complementation necessary to achieve *trans* cleavage (dashed arrows).[25] The model also highlights the dramatic conformational rearrangements required to achieve *cis* cleavage (solid arrows) and reverse the sense of the intersubunit interactions. That the C170 crystal structure was consistent with either *cis* or *trans* cleavage has already been noted.[11]

It should be stressed that because we sought to prevent assembly of higher order complexes that would further complicate NMR analysis, none of the experiments described here were performed on intact (full) attachment sites. Thus, the NMR data presented here do not allow us to directly test the tetrameric model. However, we were led to build and examine the tetrameric model because our data on the half-site complex were inconsistent with the unequivocally observed substrate cleavage and formation of suicide covalent complexes, and the resistance of this complex to proteolysis.

Although the tetrameric model provides a seductive picture of how *trans* substrate cleavage could be accommodated by λ-Int, it is important to consider whether such a model is consistent with the available biochemical data and either the *cis* or *trans* mechanisms. Experimental evidence in favor of the *trans* cleavage pathway in λ-Int was provided by complementation experiments in which two differentially cleavage-deficient λ-Int mutants, carrying mutations in one of the core residues (R212Q), and the tyrosine nucleophile (Y342F).[25] However, evidence has been contested by reports on λ-Int and Cre that indicate that such complementation experiments are subject to various artifacts.[27,59,60] Furthermore, evidence in favor of *cis* cleavage in λ-Int comes from experiments in which the differential DNA binding specificity of λ-Int and the related HK022 integrase was exploited to show that λ-Int could not rescue HK022 Int-Y/F preincubated cleavage-deficient complexes in trans. In addition, both integrases can cleave half-site substrates efficiently, even when conditions are not optimal for DNA-mediated dimerization.[22,26,27]

## Conclusion

We sought to characterize the functional role of dynamics and conformational change in the catalytic domain of λ-Int. We found that: (1) although the C terminus of the protein is indeed

flexible in solution, a conformational transition required for the *cis* cleavage mechanism appears not to occur upon 1:1 binding to a half-site substrate, (2) the accessibility of the C terminus of the protein in covalent and non-covalent complexes with DNA is clearly different, and (3) the *trans* cleavage mechanism seems to be easily accommodated by the docking models. However, we should emphasize that the data presented here are insufficient to argue in favor of either the *cis* or *trans* mechanism. For instance, it is possible that the *closed* conformation needed for *cis* cleavage is sampled in C170, at a population too low to be detected ($< \sim 5\%$). In addition, it is possible that important protein–DNA interactions in λ-Int that are needed to push the equilibrium of the C-terminal residues from the *open* to the *closed* state are insufficiently stabilized in this weak C170/DNA complex. For instance, we find that adding the (recombinantly expressed and purified) central domain comprising residues 65 to 169 of λ-Int to the reaction mixture strongly enhances DNA cleavage by C170 (unpublished results).

Is the function of the C terminus linked to its flexibility? Clearly, substantial flexibility would be necessary to accommodate the conformation that would enable *cis* cleavage. By controlling the conformational transition of the C-terminal hairpin from *open* to *closed* states *via* such allosteric means would indeed ensure a very high level of control over DNA cleavage and site-specific recombination. Conversely, as argued for Flp,[31,32] flexibility of the residues in the inter-protomer interface, and the region containing the tyrosine nucleophile in particular, could be required in order to facilitate the central isomerization that regulates the directionality of recombination and modulates protomer activity.

It is clear that characterizing the conformation and/or dynamics of the C terminus of λ-Int in covalent and non-covalent complexes with half-site and full-site DNA substrates will be critical to understanding the function of the enzyme. In the absence of high-resolution experimentally determined structures of the complexes, the models presented here should facilitate the design of biochemical experiments to probe the mechanism of λ-Int function.

## Materials and Methods

### Protein preparation

The DNA plasmid encoding C170, the catalytic domain of λ-integrase, under control of a T7 promoter, and carrying the gene for ampicillin resistance,[16,17] was provided by Arthur Landy (Brown University). The C170 gene contains eight instances of the rare AGA/AGG arginine codons, three of which occur in the first ten amino acids. The frequency of occurrence of these codons implied that overexpression would be highly dependent on the availability of the corresponding rare tRNAs.[61,62] Consequently, the *argU* gene product (also known as *dnaY*) encoding an arginyl tRNA that can decode these rare codons,[61,63] expressed on a vector encoding kanamycin resistance (pARG-U), was co-transformed into C170-expressing cells to boost protein production. C170 was overexpressed in doubly-transformed BL21(DE3) cells grown in LB at 37 °C in the presence of 50 mg/l carbenicillin and 30 mg/l kanamycin. Cultures were induced with 1 mM IPTG at an $OD_{600}$ of 0.8 and were grown for six hours after induction. The cells were harvested by centrifugation (4000*g*, 10 minutes, 4 °C) and stored at −20 °C until lysed. C170 was purified from cell pellets as previously described[16] except that the last step in the purification protocol (a hydroxylapatite column) was replaced with a size exclusion column (TosoHaas G2000SW, 2.15 × 30 cm, 3 ml/min) in buffer A (50 mM sodium phosphate pH 6.3, 0.1 mM EDTA, 1 mM DTT and 100 mM NaCl). Sample purity was judged to be >95% by electrospray mass spectrometry and SDS-PAGE. All buffers were extensively degassed with argon and fresh DTT stock solutions were added to the buffers immediately before use.

Uniformly $^{15}N$ labeled samples of C170 were obtained by growing the bacteria in M9 minimal medium[64] enriched with vitamins (Basal Eagle Vitamin Mix, Gibco), with [$^{15}N$]-$NH_4Cl$ as the sole nitrogen source; uniformly $^{15}N/^{13}C$-labeled samples were obtained by using 2 g/l [*U*-$^{13}C_6$]-glucose (Martek or CIL) as the sole carbon source. NMR samples were 180–500 μM C170 in 90% $H_2O$, 10% $D_2O$, buffer A containing 0.2% $NaN_3$, 1 μg/ml aprotinin (Pharmacia) and 1 μg/ml pepstatin (Pharmacia). NMR buffers were extensively degassed with argon, and samples sealed with Parafilm.

The protein eluted at the time expected for a monomer from a size exclusion column. Solution conditions for NMR studies were identified by screening buffer, pH and salt *via* microdialysis[65] while seeking to minimize protein precipitation. A buffer containing 50 mM phosphate (pH 6.3), 100 mM NaCl with 1 mM DTT was selected for NMR experiments although precipitation was not eliminated under these conditions. Protein samples were stable if stored at 4 °C and low concentration (<100 μM). Samples for NMR could be concentrated to ~500 μM, but at room temperature these tended to partially precipitate with a half-life of roughly three days. The precipitate was visible, but other than affecting the intensity of signals observed in NMR spectra, did not result in changes in peak positions or shapes, or appearance of new signals. For each series of NMR experiments, multiple [*U*-$^{15}N$]- and/or [*U*-$^{13}C/^{15}N$]-C170 samples were prepared such that a minimum of time elapsed between lysis of frozen bacterial cell pellets and recording of NMR spectra (usually ~18–48 hours). Even with such uninterrupted sample preparation, each sample could be used for only 2–3 days before precipitation had significantly degraded signal-to-noise, requiring use of fresh samples.

### DNA oligonucleotides

The DNA oligonucleotides were obtained in purified and desalted form (MWG Biotech). The sequences of the DNA substrates used in these experiments are shown in Figure 3(b). The DNA substrates employed contained either CTTG[40] or GAA[39] hairpin turns and were engineered to optimize stability and facilitate NOE-based DNA sequential assignments.[66] The DNA hairpins were

assembled by mixing equimolar concentrations of the DNA strands dissolved either in buffer B (10 mM sodium phosphate pH 6.3, 50 mM NaCl, 0.1 mM EDTA) or deionized water (18.2 Ω MilliQ, Millipore, Inc), heating it to 95 °C for five minutes and quenching it in ice for 15 minutes, and were confirmed by recording concentration dependent UV melts.[40] For experiments designed to monitor formation of the covalent protein–DNA complex, the top strand of oligonucleotide, d(CCC TGCCAACTTTTTT), was radiolabeled with $^{32}$P at the 5′-end by standard methods[64] by use of [γ-$^{32}$P]ATP and T4 polynucleotide kinase. The bottom strand of the *C*′s oligo was chemically phosphorylated at the 5′ end (MWG Biotech). The identity and purity of all oligos was established by negative ion MALDI-MS (MWG Biotech).

### Preparation of covalent and non-covalent protein–DNA complexes

Limited proteolysis experiments were performed on covalent and non-covalent complexes between C170 and the suicide *C*′s DNA substrate (Figure 3(b)). Non-covalent complexes were assembled by incubating protein and DNA (10 μM or 180 μM) in either equimolar concentrations or slight DNA excess in buffer A for 15 minutes. The covalent complexes were obtained by incubating an excess of C170 (10 μM) with radiolabeled oligonuclotide *C*′s (0.1 μM) for ~60 hours in buffer C (25 mM Tris (pH 7.65), 100 mM NaCl, 5 mM EDTA, 5 mM DTT);[4] formation of the suicide complex was monitored by electrophoretic mobility shift using 18% SDS-PAGE gels and visualized by autoradiography.

The C170/DNA complexes for NMR experiments were assembled by titrating DNA into $^{15}$N labeled protein. To determine whether binding and dissociation occurred in the fast exchange regime, and whether binding was complete under the conditions of the experiment, formation of the complex was monitored by recording HSQC spectra at 0, 0.5, 1 and 1.5 equivalents of DNA:protein. The spectrum of the C170/*B*′s 1:1 complex (Figure 6) was recorded at a final concentration of 187 μM. No covalent complex was detected by mass spectrometric analysis of the NMR samples.

### Limited proteolysis

The limited trypsin proteolysis experiments were carried out at room temperature in buffer C with 5% (w/w) trypsin (Sigma) and quenched at various times with 5 mM PMSF (phenyl-methylsulfonyl fluoride; Boehringer Manheim). The proteolysis products of free C170 were analyzed by positive ion electrospray mass spectrometry and by 4–20% SDS-PAGE (Biorad) visualized by Coomassie blue staining (Pierce). The proteolysis products of the covalent complexes were run on (denaturing) 18% SDS-PAGE gels and visualized by autoradiography.

### NMR spectroscopy

NMR experiments were performed at 298 K on Bruker DRX-800, DRX-600 or DMX-600 spectrometers using water flip-back[67] and gradient coherence selection.[68,69] Two-dimensional $^{15}$N-HSQC spectra of free C170 and the C170–DNA complexes were recorded at 800 MHz ($^1$H) in buffer A at concentrations of 180–500 μM.

#### *Resonance assignments*

Partial backbone resonance assignments were obtained from careful analysis of 3D HNCA,[70] HN(CO)CA,[71] HNCACB,[72] CBCA(CO)NH,[73] $^{15}$N-separated TOCSY (DIPSI-2, 8.3 kHz nominal bandwidth, $\tau_m = 60$ ms) and $^{15}$N-separated NOESY ($\tau_m = 200$ ms) spectra. In particular, many of the assigned fragments were obtained by taking advantage of the readily distinguished $C^\alpha$ shifts of the glycine, serine and threonine residues to gain inroads into the upstream and downstream residue assignments.

Identifying the three side-chain indole NH signals to the resonances shown in Figure 6(b) was achieved on the basis of $^{15}$N and $^1$H chemical shifts,[44] the absence of correlations to these amides in high-sensitivity HNCO and HNCA spectra and from characteristic NOEs to $H^{\delta 1}$ and $H^{\zeta 33}$ resonances observed in the $^{15}$N-separated NOESY spectrum. Sequence-specific assignment of the indole signals was reached by the following additional considerations. First, in the crystal structure of C170 (Figure 2(a))[11] the side-chain of Trp$^{198}$ is exposed to solvent and is positioned near the recognition helix E, while the side-chain of Trp$^{221}$ is buried in an interaction with helix D where it is buried by the side-chain of Lys$^{263}$ and makes a hydrogen bond with the backbone carbonyl of Gly$^{268}$; in the H/D exchange experiments, only one of the indole NH signals persisted into the dead-time of the exchange experiment (one hour). This observation makes that indole a likely candidate for Trp$^{221}$. Inspection of $^{15}$N-separated NOESY spectra revealed weak NOEs between the backbone amide proton of Gly$^{268}$ and the indole $H^{\varepsilon 1}$ and $H^{\delta 1}$ protons (9.45, 7.21 ppm, respectively) of Trp$^{221}$. Second, based on unambiguous backbone resonance assignments of residues in the C terminus, Trp$^{350}$ is located on a flexible region of the protein. Measurement of $^{15}$N relaxation ($T_1$, $T_2$, *NOE*) showed that one indole was significantly more mobile than the other two (Figure 5), making it a likely candidate for Trp$^{350}$. Third, only one of the tryptophan indole signals was perturbed by formation of the protein–DNA complex (Figure 6). In the crystal structures of the Cre and Flp bound to DNA, the residue at this position (Arg$^{159}$ in Cre and Tyr$^{176}$ in Flp) mediates a contact to the DNA phosphate backbone (see Figure 7). Thus, the three tryptophan side-chain indole resonances were assigned as indicated in Figure 6.

#### *Relaxation*

$^{15}$N $T_1$, $T_2$ and {$^1$H}$^{15}$N *NOE* measurements with [*U*-$^{15}$N]-C170 were carried out at 298 K on a Bruker DRX-800 spectrometer using established pulse schemes that employ water flip-back pulses and gradient coherence selection.[74] The $^1$H carrier was positioned on the water at 4.75 ppm and the $^{15}$N carrier at 118 ppm. The spectral widths were 12,500 Hz ($^1$H) and 3816.8 or 3000 Hz ($^{15}$N), recorded as a 2048 × 64 matrix. $T_1$ and $T_2$ data were acquired with a 2 s relaxation delay between scans. $T_1$ data were acquired with inversion-recovery delays of: 120, 240, 10, 1200, 360, 40, 750, 2400 ms. $T_2$ measurements were recorded with relaxation times of 58, 18, 74, 6, 130, 42, 26, 150, 106, 10 ms. Relaxation rates were obtained from a two-parameter exponential fit to the peak intensities at each time point. The standard

deviation of the baseline spectral noise was taken to be the uncertainty in peak heights. The {$^1$H}–$^{15}$N NOE data were obtained by recording in an interleaved manner spectra with (NOE) and without (no NOE) $^1$H saturation achieved *via* a train of 120° pulses (16 kHz bandwidth) applied every 18 ms for three seconds. A net relaxation delay of five seconds was employed between the repetition of the pulse sequence either with or without proton saturation. The NOE values were determined from the ratios of the peak intensities between the two spectra. All 2D spectra were processed by applying a low-pass solvent filter, extended once by forward–backward linear prediction in the indirect dimension, apodized in both dimensions with a cosine function and zero filled to a final size of 2048 × 256. Data processing was performed with NMRPipe[75] and analyzed with NMRView.[76] Effective correlation times $\tau_c$ were estimated from the $T_1/T_2$ ratios[46,47] with the program r2r1_tm (A. G. Palmer, 1998).

Hydrogen/deuterium exchange experiments were performed on a uniformly $^{15}$N labeled sample which was exchanged into buffer A containing 100% $D_2O$ with a small gel-filtration column (PD-10, Pharmacia). Exchange was monitored at 25 °C by recording a series of 2D HSQC spectra over the course of 24 hours.

#### Molecular microbiology

Models of the complex between C170 and half-site or Holliday junction DNA substrates were generated using the coordinates of C170-Y342F in the crystal structure (chain A in 1AE9) and the Cre/*loxP* complex (2CRX). Briefly, the tetrameric Cre/*loxP* coordinates were generated from the crystallographic symmetry operations using SPDBV†.[77] Then, the $C^\alpha$ carbons of the catalytic residues of C170 ($Arg^{212}$, $His^{308}$ and $Arg^{311}$) were superposed onto the $C^\alpha$ carbons of the corresponding residues of Cre ($Arg^{173}$, $His^{298}$ and $Arg^{302}$) using MOLMOL.[78] The half-site model results from superposition of C170 onto chain A of 2CRX, while the tetrameric model was generated using chains A and B and their crystallographically related molecules. No other restraints or assumptions were imposed in the docking/modeling process. The same method was followed to dock C170 onto the Flp coordinates using the $C^\alpha$ atoms of $Arg^{190}$, $His^{305}$ and $Arg^{308}$ (not shown). Molecular renderings were generated with MOLMOL.

Hydrodynamic modeling and calculations of diffusion anisotropy and predicted relaxation rates were performed on the unmodified coordinates of 1AE9A with HYDRONMR,[49] with a bead size of 3.1 Å, viscosity of 0.01 poise, and an $^{15}$N CSA of −170 ppm.[79]

## Acknowledgements

The authors thank A. Landy, R. Tirumalai (Brown University) and G. Van Duyne (Univ. Pennsylvania) for reagents and invaluable advice, assistance and stimulating discussions, V. Gopalan for expert advice and for enabling the radioisotope experiments at OSU, H. Kamadurai, C. Amero and other members of the Foster lab for general help and stimulating discussions, P. E. Wright and J. Chung (The Scripps Research Institute) for NMR pulse programs, I.-J. Byeon (OSU) for assistance with NMR data collection, A. Palmer (Columbia University), J. Garcia de la Torre (Universidad de Murcia, Spain), F. Delaglio (NIH), B. Johnson (Merck) and K. Wüthrich (ETH, Switzerland) for software, and the OSU CCIC for NMR and MS instrumentation services. This work was supported in part by grants to MF from the Ohio State Research Foundation, the ACS/Petroleum Research Fund (34006-G4) and the National Science Foundation (MCB0092962), and by grants to A. Landy from the National Institutes of Health (GM33928 and GM62723).

† http://www.expasy.org/spdbv

## References

1. Craig, N. L. & Nash, H. A. (1983). The mechanism of phage lambda site-specific recombination: site-specific breakage of DNA by Int topoisomerase. *Cell*, **35**, 795–803.
2. Echols, H., Dodson, M., Better, M., Roberts, J. D. & McMacken, R. (1984). The role of specialized nucleoprotein structures in site-specific recombination and initiation of DNA replication. *Cold Spring Harb. Symp. Quant. Biol.* **49**, 727–733.
3. Richet, E., Abcarian, P. & Nash, H. A. (1986). The interaction of recombination proteins with supercoiled DNA: defining the role of supercoiling in lambda integrative recombination. *Cell*, **46**, 1011–1021.
4. Nunes-Duby, S. E., Matsumoto, L. & Landy, A. (1987). Site-specific recombination intermediates trapped with suicide substrates. *Cell*, **50**, 779–788.
5. Landy, A. (1989). Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu. Rev. Biochem.* **58**, 913–949.
6. Stark, W. M., Boocock, M. R. & Sherratt, D. J. (1992). Catalysis by site-specific recombinases. *Trends Genet.* **8**, 432–439.
7. Nash, H. A. (1996). *Site-specific recombination: integration, excision, resolution, and inversion of defined DNA segments Escherichia coli* and *Salmonella*: Cellular and Molecular Biology (Neidhardt, F. C. & Curtiss, R., eds), 2nd edit., pp. 2363–2376, ASM Press, Washington, DC.
8. Azaro, M. A. & Landy, A. (2002). *Lambda integrase and the lambda Int family* Mobile DNA II (Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M., eds), ASM Press, Washington, DC.
9. Nunes-Duby, S. E., Azaro, M. A. & Landy, A. (1995). Swapping DNA strands and sensing homology without branch migration in lambda site-specific recombination. *Curr. Biol.* **5**, 139–148.
10. Azaro, M. A. & Landy, A. (1997). The isomeric preference of Holliday junctions influences resolution bias by lambda integrase. *EMBO J.* **16**, 3744–3755.
11. Kwon, H. J., Tirumalai, R., Landy, A. & Ellenberger, T. (1997). Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science*, **276**, 126–131.
12. Van Duyne, G. D. (2001). A structural view of cre-loxp site-specific recombination. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 87–104.

13. Moitoso de Vargas, L., Pargellis, C. A., Hasan, N. M., Bushman, E. W. & Landy, A. (1988). Autonomous DNA binding domains of lambda integrase recognize two different sequence families. *Cell*, **54**, 923–929.
14. Han, Y. W., Gumport, R. I. & Gardner, J. F. (1994). Mapping the functional domains of bacteriophage lambda integrase protein. *J. Mol. Biol.* **235**, 908–925.
15. Sarkar, D., Radman-Livaja, M. & Landy, A. (2001). The small DNA binding domain of lambda integrase is a context-sensitive modulator of recombinase functions. *EMBO J.* **20**, 1203–1212.
16. Tirumalai, R. S., Kwon, H. J., Cardente, E. H., Ellenberger, T. & Landy, A. (1998). Recognition of core-type DNA sites by lambda integrase. *J. Mol. Biol.* **279**, 513–527.
17. Tirumalai, R. S., Healey, E. & Landy, A. (1997). The catalytic domain of lambda site-specific recombinase. *Proc. Natl Acad. Sci. USA*, **94**, 6104–6109.
18. Pargellis, C. A., Nunes-Duby, S. E., de Vargas, L. M. & Landy, A. (1988). Suicide recombination substrates yield covalent lambda integrase–DNA complexes and lead to identification of the active site tyrosine. *J. Biol. Chem.* **263**, 7678–7685.
19. Esposito, D. & Scocca, J. J. (1997). The integrase family of tyrosine recombinases: evolution of a conserved active site domain. *Nucl. Acids Res.* **25**, 3605–3614.
20. Nunes-Duby, S. E., Kwon, H. J., Tirumalai, R. S., Ellenberger, T. & Landy, A. (1998). Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucl. Acids Res.* **26**, 391–406.
21. Cheng, C., Kussie, P., Pavletich, N. & Shuman, S. (1998). Conservation of structure and mechanism between eukaryotic topoisomerase I and site-specific recombinases. *Cell*, **92**, 841–850.
22. Tekle, M., Warren, D. J., Biswas, T., Ellenberger, T., Landy, A. & Nunes-Duby, S. E. (2002). Attenuating functions of the C-terminus of lambda integrase. *J. Mol. Biol.* **324**, 649–665.
23. Mondragon, A. (1997). Solving the *cis/trans* paradox in the Int family of recombinases. *Nature Struct. Biol.* **4**, 427–429.
24. Grindley, N. D. (1997). Site-specific recombination: synapsis and strand exchange revealed. *Curr. Biol.* **7**, R608–R612.
25. Han, Y. W., Gumport, R. I. & Gardner, J. F. (1993). Complementation of bacteriophage lambda integrase mutants: evidence for an intersubunit active site. *EMBO J.* **12**, 4577–4584.
26. Nunes-Duby, S. E., Tirumalai, R. S., Dorgai, L., Yagil, E., Weisberg, R. A. & Landy, A. (1994). Lambda integrase cleaves DNA in cis. *EMBO J.* **13**, 4421–4430.
27. Nunes-Duby, S. E., Radman-Livaja, M., Kuimelis, R. G., Pearline, R. V., McLaughlin, L. W. & Landy, A. (2002). Lambda integrase complementation at the level of DNA binding and complex formation. *J. Bacteriol.* **184**, 1385–1394.
28. Gopaul, D. N., Guo, F. & Van Duyne, G. D. (1998). Structure of the Holliday junction intermediate in Cre-loxP site- specific recombination. *EMBO J.* **17**, 4175–4187.
29. Guo, F., Gopaul, D. N. & van Duyne, G. D. (1997). Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature*, **389**, 40–46.
30. Guo, F., Gopaul, D. N. & Van Duyne, G. D. (1999). Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse. *Proc. Natl Acad. Sci. USA*, **96**, 7143–7148.
31. Chen, Y., Narendra, U., Iype, L. E., Cox, M. M. & Rice, P. A. (2000). Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping. *Mol. Cell*, **6**, 885–897.
32. Conway, A. B., Chen, Y. & Rice, P. A. (2003). Structural plasticity of the flp-holliday junction complex. *J. Mol. Biol.* **326**, 425–434.
33. Chen, J. W., Lee, J. & Jayaram, M. (1992). DNA cleavage in trans by the active site tyrosine during Flp recombination: switching protein partners before exchanging strands. *Cell*, **69**, 647–658.
34. Lee, J., Whang, I. & Jayaram, M. (1994). Directed protein replacement in recombination full sites reveals trans-horizontal DNA cleavage by Flp recombinase. *EMBO J.* **13**, 5346–5354.
35. Dixon, J. E., Shaikh, A. C. & Sadowski, P. D. (1995). The Flp recombinase cleaves Holliday junctions in trans. *Mol. Microbiol.* **18**, 449–458.
36. Nunes-Duby, S. E., Matsumoto, L. & Landy, A. (1989). Half-att site substrates reveal the homology independence and minimal protein requirements for productive synapsis in lambda excisive recombination. *Cell*, **59**, 197–206.
37. Nunes-Duby, S. E., Yu, D. & Landy, A. (1997). Sensing homology at the strand-swapping step in lambda excisive recombination. *J. Mol. Biol.* **272**, 493–508.
38. Knudsen, B. R., Dahlstrom, K., Westergaard, O. & Jayaram, M. (1997). The yeast site-specific recombinase Flp mediates alcoholysis and hydrolysis of the strand cleavage product: mimicking the strand-joining reaction with non-DNA nucleophiles. *J. Mol. Biol.* **266**, 93–107.
39. Hirao, I., Kawai, G., Yoshizawa, S., Nishimura, Y., Ishido, Y., Watanabe, K. & Miura, K. (1994). Most compact hairpin-turn structure exerted by a short DNA fragment, d(GCGAAGC) in solution: an extraordinarily stable structure resistant to nucleases and heat. *Nucl. Acids Res.* **22**, 576–582.
40. Blommers, M. J., Walters, J. A., Haasnoot, C. A., Aelen, J. M., van der Marel, G. A., van Boom, J. H. & Hilbers, C. W. (1989). Effects of base sequence on the loop folding in DNA hairpins. *Biochemistry*, **28**, 7491–7498.
41. van Dongen, M. J., Mooren, M. M., Willems, E. F., van der Marel, G. A., van Boom, J. H., Wijmenga, S. S. & Hilbers, C. W. (1997). Structural features of the DNA hairpin d(ATCCTA-GTTA-TAGGAT): formation of a G-A base pair in the loop. *Nucl. Acids Res.* **25**, 1537–1547.
42. Erie, D., Sinha, N., Olson, W., Jones, R. & Breslauer, K. (1987). A dumbbell-shaped, double-hairpin structure of DNA: a thermodynamic investigation. *Biochemistry*, **26**, 7150–7159.
43. Kwon, H. J. (1999). Structural Studies of Lambda Integrase and the Crystal Structure of the RobA Transcription Factor. PhD, Harvard University.
44. Seavey, B. R., Farr, E. A., Westler, W. M. & Markley, J. L. (1991). A relational database for sequence-specific protein NMR data. *J. Biomol. NMR*, **1**, 217–236.
45. Grzesiek, S. & Bax, A. (1993). Amino acid type determination in the sequential assignment procedure of

uniformly $^{13}C/^{15}N$-enriched proteins. *J. Biomol. NMR*, **3**, 185–204.
46. Kay, L. E., Torchia, D. A. & Bax, A. (1989). Backbone dynamics of proteins as studied by $^{15}N$ inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, **28**, 8972–8979.
47. Tjandra, N., Feller, S. E., Pastor, R. W. & Bax, A. (1995). Rotational diffusion anisotropy of human ubiquitin from $^{15}N$ NMR relaxation. *J. Am. Chem. Soc.* **117**, 12562–12566.
48. Kamadurai, H. B., Subramaniam, S., Jones, R. B., Green-Church, K. B. & Foster, M. P. (2003). Protein folding coupled to DNA binding in the catalytic domain of bacteriophage lambda integrase detected by mass spectrometry. *Protein Sci.* **12**, 620–626.
49. Garcia de la Torre, J., Huertas, M. L. & Carrasco, B. (2000). HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Reson.* **147**, 138–146.
50. Foster, M. P., Wuttke, D. S., Clemens, K. C., Jahnke, W., Radhakrishnan, I., Tennant, L. *et al.* (1998). Chemical shift as a probe of molecular interfaces: NMR studies of DNA binding by the three amino-terminal zinc finger domains from transcription factor IIIA. *J. Biomol. NMR*, **12**, 51–71.
51. Blakely, G. & Sherratt, D. (1996). Determinants of selectivity in Xer site-specific recombination. *Genes Dev.* **10**, 762–773.
52. Subramanya, H. S., Arciszewska, L. K., Baker, R. A., Bird, L. E., Sherratt, D. J. & Wigley, D. B. (1997). Crystal structure of the site-specific recombinase, XerD. *EMBO J.* **16**, 5178–5187.
53. Grishin, N. V. (2000). Two tricks in one bundle: helix-turn-helix gains enzymatic activity. *Nucl. Acids Res.* **28**, 2229–2233.
54. Ross, W. & Landy, A. (1983). Patterns of lambda Int recognition in the regions of strand exchange. *Cell*, **33**, 261–272.
55. Hoess, R., Abremski, K., Irwin, S., Kendall, M. & Mack, A. (1990). DNA specificity of the Cre recombinase resides in the 25 kDa carboxyl domain of the protein. *J. Mol. Biol.* **216**, 873–882.
56. Bruckner, R. C. & Cox, M. M. (1986). Specific contacts between the FLP protein of the yeast 2-micron plasmid and its recombination site. *J. Biol. Chem.* **261**, 11798–11807.
57. Martin, S. S., Pulido, E., Chu, V. C., Lechner, T. S. & Baldwin, E. P. (2002). The order of strand exchanges in Cre-LoxP recombination and its basis suggested by the crystal structure of a Cre-LoxP Holliday junction complex. *J. Mol. Biol.* **319**, 107–127.
58. Kazmierczak, R. A., Swalla, B. M., Burgin, A. B., Gumport, R. I. & Gardner, J. F. (2002). Regulation of site-specific recombination by the C-terminus of lambda integrase. *Nucl. Acids Res.* **30**, 5193–5204.
59. Shaikh, A. C. & Sadowski, P. D. (1997). The Cre recombinase cleaves the lox site *in trans*. *J. Biol. Chem.* **272**, 5695–5702.
60. Shaikh, A. C. & Sadowski, P. D. (2000). Chimeras of the Flp and Cre recombinases: tests of the mode of cleavage by Flp and Cre. *J. Mol. Biol.* **302**, 27–48.
61. Brinkmann, U., Mattes, R. E. & Buckel, P. (1989). High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the dnaY gene product. *Gene*, **85**, 109–114.
62. Zahn, K. & Landy, A. (1996). Modulation of lambda integrase synthesis by rare arginine tRNA. *Mol. Microbiol.* **21**, 69–76.
63. Saxena, P. & Walker, J. R. (1992). Expression of argU, the *Escherichia coli* gene coding for a rare arginine tRNA. *J. Bacteriol.* **174**, 1956–1964.
64. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2 edit., Cold Spring Harbor, NY.
65. Bagby, S., Tong, K. I., Liu, D., Alattia, J. R. & Ikura, M. (1997). The button test: a small scale method using microdialysis cells for assessing protein solubility at concentrations suitable for NMR. *J. Biomol. NMR*, **10**, 279–282.
66. Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, Wiley, New York.
67. Grzesiek, S. & Bax, A. (1993). The importance of not saturating $H_2O$ in protein NMR. Application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* **115**, 12593–12594.
68. Muhandiram, D. R. & Kay, L. E. (1994). Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson. Ser. B*, **103**, 203–216.
69. Sattler, M., Schleucher, J. & Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progr. Nucl. Magn. Reson. Spectrosc.* **34**, 93–158.
70. Clore, G. M., Bax, A., Driscoll, P. C., Wingfield, P. T. & Gronenborn, A. M. (1990). Assignment of the side-chain $^{1}H$ and $^{13}C$ resonances of interleukin-1 beta using double- and triple-resonance heteronuclear three-dimensional NMR spectroscopy. *Biochemistry*, **29**, 8172–8184.
71. Bax, A. & Ikura, M. (1991). An efficient 3D NMR technique for correlating the proton and $^{15}N$ backbone amide resonances with the alpha-carbon of the preceding residue in uniformly $^{15}N/^{13}C$ enriched proteins. *J. Biomol. NMR*, **1**, 99–104.
72. Wittekind, M. & Mueller, L. (1993). HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *J. Magn. Reson.* B**101**, 201–205.
73. Grzesiek, S. & Bax, A. (1993). Correlating backbone amide and side-chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114**, 6291–6293.
74. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G. *et al.* (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by $^{15}N$ NMR relaxation. *Biochemistry*, **33**, 5984–6003.
75. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
76. Johnson, B. A. & Blevins, R. A. (1994). NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603–614.
77. Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.
78. Koradi, R., Billeter, M. & Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55. see also pp. 29–32.
79. Tjandra, N., Szabo, A. & Bax, A. (1996). Protein

backbone dynamics and $^{15}N$ chemical shift anisotropy from quantitative measurement of relaxation interference effects. *J. Am. Chem. Soc.* **118**, 6986–6991.

80. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305–308.

***Edited by M. F. Summers***

*(Received 2 December 2002; received in revised form 2 April 2003; accepted 2 April 2003)*

*Note added in proof*: Recent work in the laboratories of A. Landy (Brown University) and T. Ellenberger (Harvard University), has enabled crystallization of a construct containing the central and catalytic domains of λ-Int in a *cis* covalent complex with a DNA substrate (A. Landy and S. Nunes-Duby, personal communication). Their forthcoming report provides what may be the most definitive evidence to date that λ-Int cleaves its DNA substrates in *cis*, not *trans*.