Cre Mutants with Altered DNA Binding Properties*

Markus Hartung and Brigitte Kisters-Woike‡

From the Institute for Genetics, University of Cologne, Weyertal 121, D-50931 Cologne, Germany

The recombinase Cre of bacteriophage P1 is a member of the family of site-specific recombinases and integrases that catalyze inter- and intramolecular DNA rearrangements. To understand how this protein specifically recognizes its target sequence, we constructed Cre mutants with amino acid substitutions in different positions of the presumptive DNA binding region. Here we present the results of in vitro DNA binding and in vivo recombination experiments with these Cre mutants. Most substitutions of presumptive DNA-binding amino acids in in vitro tests resulted either in the loss of target binding or in a broadening of target recognition specificity. Of the mutations resulting in a broadening of target specificity, one, N317A, results in a reduced recombination efficiency with the wild-type loxP target but recombines, in contrast to wild-type Cre, in in vivo experiments, with a symmetric variant of the wild-type target sequence. This target variant differs from wild-type loxP by the symmetric C to A replacement in position 6 of the inverted repeats. We propose a common multithetical DNA binding motif for the family of integrases and recombinases. This model implies a major structural rearrangement for the DNA binding region of integrase, analogous to the structural rearrangements of the DNA binding motifs of other proteins when contacting their target DNA.

In site-specific recombination, DNA molecules are cleaved in both strands at two separate recombination sites, and the ends are rejoined to new partners (1). The reaction is carried out without any synthesis or degradation of DNA. Two families of site-specific recombinases are known, the resolvase/invertase family and the λ integrase family. In site-specific recombinations mediated by both families, four recombinases bound to two target sites catalyze the reaction (2). Each target site consists of a core site of approximately 30 base pairs. The core site can be characterized as two inverted repeats, separated by a spacer region. The spacer regions differ in length, depending on the respective protein. The resolvase/invertase family uses a serine nucleophile to mediate a concerted double strand cleavage. The rejoining in this case occurs at nucleotide phosphates separated by 2 base pairs. Contrary to the resolvases/invertases, the integrases use a tyrosine nucleophile to mediate sequential strand exchange. The cleavage of the strands occurs in a staggered manner, at nucleotides that are 6–8 base pairs apart.

The family of λ integrases was first described with respect to their amino acid sequences by Argos et al. (3). The alignment of the C-terminal domains of these recombinases revealed a strictly conserved motif consisting of three amino acids as follows: His (2 amino acids), Arg (31 amino acids), and Tyr. A conserved arginine, located more N-terminally was found later in addition to the “conserved triad” (4). It was proposed that this four-amino acid motif, which is common to all known integrases, is responsible for DNA cleavage.

The Cre-lox site-specific recombination system of bacteriophage P1 was first described in 1978 by Sternberg and co-workers (5, 6). Like other site-specific recombination systems it consists of the following two components: the recombinase itself, Cre (causes recombination (7)), and its target sequence, loxP (locus of crossing over (8)), P1 (7). Cre recombinase is a member of the λ integrase family. This protein is a single polypeptide chain of 343 amino acids (38.5 kDa) (8). In solution it exists as a monomer (9), and no additional proteins are needed to carry out efficient recombination (10).

Limited proteolysis of Cre by chymotrypsin generates two small fragments of approximately 13.5 kDa and a large 25-kDa fragment (11). The 25-kDa fragment contains the C terminus of Cre and is able to bind to the lox site (7) but only with reduced affinity as compared with the full size protein. Sequence analysis localized the point of crossing over in the phage DNA to an 8-bp spacer region between two perfect 13-bp inverted repeats, the loxP site (12). This suggested that the inverted repeats serve as recognition elements for the binding of Cre to DNA and that each loxP site is bound by two Cre molecules (13). Length and orientation of the spacer region are of importance for the recombination process. Each Cre cuts behind the first base of the spacer region on each side, leaving 5'-overhanging ends of 6 base pairs (14). As with the other integrases (3, 4), the cleavage results in a covalent attachment of the conserved Tyr-324 of Cre to a 3'-PO4. Next, the physical exchange of DNA strands takes place. Furthermore, the spacer region defines the orientation of the two lox sites that are involved in the recombination event: depending on the relative orientation of the two sites to each other, the DNA between the two sites is either excised or inverted. Excision occurs when the spacer regions of the two lox sites are directly repeated. Inversion takes place when they are in opposite orientations (15).

Cre-mediated recombination is not restricted to bacterial host cells. In eukaryotes it was first demonstrated by Cre-mediated deletion of a LEU2 gene flanked by two lox sites (16). Afterward, by Cre-mediated recombination, the LEU2 gene was deleted. After the awareness that Cre can efficiently catalyze recombination of lox sites in eukaryotic cells, the Cre-lox recombination system was rapidly established as a general tool for studying gene function (e.g. see Gu et al. (17)).

The question of how Cre specifically recognizes its target DNA is not solved in detail yet, although recently the structure of this recombinase bound to its target sequence has been reported to 2.4 Å resolution (18). The analysis of the crystal structure of Cre reveals an N-terminal domain, the C-terminal domain, and an extended sequence. The C-terminal domain, which encompasses the site of DNA cleavage, has a compact structure. The N-terminal domain, which is involved in target recognition, is less well ordered.

* This work was supported by the German Israel Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed. Tel.: 49-221-470-3964; Fax: 49-221-470-5170; E-mail: b.kisters@uni-koeln.de.
structure with respect to protein function focused on the active sites of the synaptic complex and on the mechanism of Cre-loxP site-specific recombination and less on the target recognition specificity of the protein (18). The crystals of the catalytic cores of HP1 and β integrase (19, 20), or the crystal of XerD (21), give no final answer to the question of how these proteins specifically recognize their target sequences because these proteins had been crystallized without their target DNA.

It has been shown that Cre induces a very modest bend in the DNA when it binds (22). The overall affinity of Cre for the lox site might be increased by this bend. Moreover, the bend might also assist the recombination process by facilitating helix opening during the exchange of DNA strands. Several mutants of Cre with reduced DNA binding affinity have been identified by gel retardation assays with DNA fragments containing a lox site as target sequence (23).

We have analyzed the specific protein-DNA recognition of the Cre-lox recombination system. First, we substituted each position of the inverted repeats of loxP symmetrically for the three other bases. These targets were used to identify the nucleotide positions within the loxP site which determine the specific DNA-protein recognition by in vitro gel retardation experiments. Then, amino acid positions of Cre that are involved in protein-DNA recognition were localized on the basis of the x-ray structures of y6 resolvase (24), β integrase (20), XerD (21), HP1 integrase (19), and the predicted secondary structure of Cre. These predictions were tested by Cre mutants in vitro as well as in vivo. In addition, we present evidence for a multihelical DNA binding motif that is common for all integrases when bound to DNA.

EXPERIMENTAL PROCEDURES

Chemicals and Enzymes

Restriction enzymes were from New England Biolabs (Bad Schwalbach, Germany) or Boehringer-Mannheim (Germany); Klenow fragment of DNA polymerase I was from Boehringer Mannheim; T4 DNA ligase was from Life Technologies, Inc. (Germany); [α-32P]deoxyribonucleotides were from Amersham Buchler (Braunschweig, Germany); chemicals for automated DNA synthesis were from Applied Biosystems (Pforstung, Germany); all other chemicals from Sigma (München, Germany) or Merck (Darmstadt, Germany). Oligonucleotides were synthesized on an Applied Biosystems 380A synthesizer.

Bacterial Strains, Plasmids, and Media

Strains

BL21(DE3), F'ompT hsdSB (r'K- m'K-; an Escherichia coli B strain) with DE3, a λ prophage carrying the T7 RNA polymerase gene (25), BMH8117: F' Δlac proAB) gyrA (NalR) thi- supE λ- (26), DH15α: F'endA hsdR17(k') m'K- supE44 thi-1 recA gryA96 relA1 e'80 ΔlacZ ΔM15 (27).

Plasmids

pET11a-Cre-utr, a derivative of pET11a (25)—The Ndel-BamHI polymerase chain reaction fragment of the wild-type cre gene of bacteriophage P1 (28) replaces the Ndel-BamHI fragment of pET11a, a part of the α10 leader sequence.

pET11a-Cre-Hel, a Derivative of pET11a-Cre-utr—pET11a-Cre-Hel is an expression plasmid for Cre mutants. This plasmid differs from pET11a-Cre-utr by silent mutations plus base substitutions resulting in the amino acid substitutions of the respective protein mutants.

pACYC-Cre—The HindIII-AvaI fragment of pACYC184 (29, 30), its tetracycline resistance gene, is replaced by the HindIII-AvaI fragment of pET11a-Cre-utr including the wild-type cre gene and lacP'.

pACYC-Cre-Hel—The HindIII-Ndel fragment of pACYC184 including the wild-type cre gene is replaced by the HindIII-Ndel fragment of pET11a-Cre-Hel, which comprises the cre gene with several silent mutations.

pEst-Set-lox—The plasmid is based on pEstS00 (31, 32). The lacY gene of pUCB2 was fused in frame to the 3' terminus of the lacZ gene via an EcoRI site of pEstS00. An insert consisting of two lox sites in direct repeat orientation, which are spaced by the HindIII-AvaI fragment of lacI followed by two ochre stop codons, was inserted into a SpeI site. No lacZ/lacY-expression was observed with this construct in the absence of Cre due to the stop codons. In the presence of Cre, both stop codons plus the HindIII-AvaI fragment are removed by recombination. This results in expression of lacZ which can be easily monitored on 5-bromo-4-chloro-3-indolyl-β-D-galactoside plates.

Media

YT, DY, and NZCYM are standard media. Antibiotic concentrations are according to Sambrook et al. (33). Minimal lactose plates: 1 liter contains 860 ml of double distilled H2O, 16 g of agar. After autoclaving 100 ml of 10× Buffer A (34), 20 ml of 20% lactose, 2 ml of 1% thiamine, 1 ml of 1 M MgSO4, 10 ml of 2% proline, 200 μl of 10 mM FeSO4, and 8.8 ml of 10 mM CaCl2 were added. CTX is composed of 1 liter of medium that contains 3.3 ml of 5-bromo-4-chloro-3-indolyl-β-D-galactoside (2% in dimethylformamide), 400 μl of chloramphenicol (100 mg/ml), and 600 μl of tetracycline (10 mg/ml) in minimal plates.

Expression of Cre Protein

Crude extracts were obtained from E. coli strain BL21(DE3) containing plasmid pET11a-Cre-wt or pET11a-Cre-Hel. The culture was induced by adding isopropyl-1-thio-β-D-galactoside to a final concentration of 1 mM. Three hours after induction the cells were harvested, resuspended in TSE buffer (20 mM Tris-HCl, pH 7.5, 1 mM Na2EDTA, 10% glycerol, w/v), 0.3 mM NaCl (9), and sonified (W220F, microtip, Heat Systems-Ultrasonic, Inc.). Cre expression was controlled by Coomassie Blue-stained SDS-polyacrylamide gel electrophoresis (35). Protein concentration in the crude extract was determined according to Bradford (36) and was typically approximately 5.5 μg of protein/μl.

Preparation of DNA Substrates

Oligonucleotides of 37 bases of either the wild-type loxP site, the palindromic wild-type (5'-CCATAACTTCGTATA ATGTA 3'), or a variant of the palindromic target characterized by a symmetric single-base substitution, were purified by electrophoresis through a 12% polyacrylamide, 8 μt urea gel. Concentration was determined by absorbance at 260 nm.

DNA Binding Reactions

Experiments testing the binding of Cre to lox sites were carried out in 5% glycerol, 50 mM Tris-HCl, pH 7.5, 32 μM bovine serum albumin, 1 mM dithiothreitol, and 30 mM NaCl. Each reaction was initiated by adding 4 μl of crude extract (~5.5 μg/μl) and 100,000 cpm labeled lox site DNA (~0.5 pmol). Following incubation at 30 °C for 30 min, 1 μl (10 ng/μl) of heparin and 15 μl of 1.5% Ficoll, 0.05% bromphenol blue, and 0.1% xylene cyanol were added to each tube (13).

Gel Mobility Shifts Assay

Immediately following the addition of heparin, Ficoll, bromphenol blue, and xylene cyanol, an approximately 20 μl of the above recombination reaction mix was loaded per lane on a 5% running non-denaturating acrylamide gel. Electrophoresis was performed in 0.5× TBE for 2.5 h at 5 V/cm. The gels were dried and autoradiographed as well with screens of PhosphorImager Fuji BAS 1000, Raytest, Straubenhardt (overnight), as with Kodak XAR film at ~70 °C for 3 days.

In Vivo Test System

The in vivo test system is based on the recombination between two lox sites integrated in pEst-Set-lox. After introducing each construct into BMH8117, competent cells (37) were transformed with pACYC-Cre or mutated pACYC-Cre-Hel plasmids and plated on CTX minimal lactose plates for approximately 3 days.

Computing

Secondary structure predictions were done by PHDsec (39–41). Modelling was done with Insight II (Biosym MSI Inc., CA) on an ESV workstation.

RESULTS

The DNA Binding Properties of Crude Extract Containing Wild-type Cre with Symmetrical loxP Sites with Single Base Exchanges in the Inverted Repeats—To analyze which positions of the inverted repeats of loxP are important for the specific attachment of Cre to its target DNA, we substituted each

2 S. Oehler, personal communication.
position of the inverted repeats symmetrically for the three other bases. Gel mobility shift assays were performed with these targets and crude extracts containing wild-type Cre. Gel mobility shift assays were performed with these targets and crude extracts containing wild-type Cre. Fig. 1 depicts, as an example for this type of analysis, the results of a gel mobility shift assay of wild-type Cre and Cre mutants with wild-type loxP and two target variants in positions 2 and 6 of the inverted repeats (Fig. 1). We did not consider the spacer region in this analysis, because it has been shown previously that symmetric substitutions in this region had no effect on binding of the protein to its target DNA (15). The stoichiometry of one Cre molecule bound per single loxP inverted repeat, i.e. single half-site (complex I), and of two Cre molecules bound to the complete, symmetric site (complex II) have previously been shown in vitro on retardation gels (13). According to these results and the pattern obtained with a loxP half-site as control, the retarded bands visible in Fig. 1 were identified as complex I and complex II. We quantified the bands corresponding to complex I and complex II by PhosphorImager analysis. Fig. 2 summarizes the results of this quantification. Both complexes show the same pattern of sensitivity concerning base substitutions in the inverted repeats of the target sequence, although the effect is more marked for complex II. For both

**FIG. 1.** A and B, band shift assays of symmetrical loxP sites with symmetrical single base substitutions in position 2 or position 6 of the inverted repeats and Cre mutants. The target variants used in the band shift assay are as follows: wt, symmetrical loxP sequence; X, control, double-stranded oligonucleotide with no sequence similarity to loxP (5'-TGTCTAGAGTCT TCAAGGTG GTCCAATTCGGTG-3'); xwt2, "half-site" lox sequence (5'-GGATAACTTCGTATATA GTTTGTT CAAAT-3'); P2, symmetrical loxP variant with a single base substitution at position 2 (T -> G); P6, symmetrical loxP variant with a single base substitution at position 6 (C -> A). The Cre variants used are indicated below. Lanes 1 and 17, no crude extract. Lanes 2 and 18, D.C, crude extract without Cre. Lanes 3–7 and 19–23, crude extract with wild-type Cre. Lanes 8–16 and 24–32, crude extract with Cre with a single amino acid replaced. Symbols at the left indicate the complexes causing the respective band. From bottom to top, hairpin; double-stranded free DNA; one molecule Cre bound to the symmetrical target; two molecules Cre bound to the symmetrical target. In each lane 2.6 ng of heparin was added as competitor. C, sequence of the symmetrical loxP site used as wild type in the band shift assays. The inverted repeats are indicated by horizontal arrows and numbered according to the positions of the nucleotides of the inverted repeats of the loxP site starting with the first nucleotide; small vertical arrows indicate the sites where the target would be cut during the recombination event; larger vertical arrows indicate the base pair substitutions in positions 2 and 6 of the inverted repeats (Fig. 1). We did not consider the spacer region in this analysis, because it has been shown previously that symmetric substitutions in this region had no effect on binding of the protein to its target DNA (13). The stoichiometry of one Cre molecule bound per single loxP inverted repeat, i.e. single half-site (complex I), and of two Cre molecules bound to the complete, symmetric site (complex II) have previously been shown in vitro on retardation gels (15). According to these results and the pattern obtained with a loxP half-site as control, the retarded bands visible in Fig. 1 were identified as complex I and complex II. We quantified the bands corresponding to complex I and complex II by PhosphorImager analysis. Fig. 2 summarizes the results of this quantification. Both complexes show the same pattern of sensitivity concerning base substitutions in the inverted repeats of the target sequence, although the effect is more marked for complex II. For both
complexes, positions 2 and 6 are the most sensitive positions for binding of Cre to its target, followed by positions 3 and 7. Positions beyond 9 seem to be of no or inferior importance for the binding of Cre to its target in vitro.

What Is the DNA Binding Motif of Cre? Design of Mutants with Altered Binding Specificity—Recently, the co-crystal of Cre bound to its target has been published (18). However, at the time we commenced this, no information concerning the structure of Cre was available. To design mutants with altered DNA binding specificity, we had to rely on the x-ray analyses of closely related proteins, the integrases λ Int, XerD, and HP1, and the x-ray data of a more distantly related protein, the γδ resolvase. The latter data had to be considered because γδ resolvase, in contrast to the closely related proteins, had been crystallized as a co-complex with its target DNA.

The family of site-specific recombinases/integrases is well defined by a conserved pattern of a Arg/His/Arg triad, accompanied by a conserved Tyr, which is covalently bound to a 3′-phosphate of the DNA site cut during the DNA rearrangement. The conserved motif is located in the C-terminal region of these proteins. Their N-terminal parts are largely divergent. In order to identify those amino acids of Cre which might be involved in DNA recognition, we first aligned the amino acid sequences of Cre, γδ resolvase, λ Int, XerD, and HP1 with respect to their secondary structure elements. Although γδ resolvase is not directly related to the integrases, we had to include this protein in the alignment because, as mentioned above, γδ resolvase was the only related protein crystallized as a co-complex bound to DNA. In order to identify the positions of those amino acids that might be involved in DNA recognition, we compared the three-dimensional structure of DNA-bound γδ resolvase to the crystal structures of the non-bound integrases.

Finally, the alignment of the secondary structures of the integrases allowed us to identify those amino acids of Cre that might be involved in the specific DNA recognition of this protein.

From these considerations, we decided to concentrate on two regions of Cre. First, we chose a region that was predicted to be mainly α-helical, surrounding the catalytic tyrosine 324. Second we chose a stretch of amino acids, also predicted to be α-helical, immediately following the catalytic residues histidine 289 and arginine 292. From these regions, several positions were chosen for amino acid substitutions. The resulting mutants were analyzed with respect to their DNA recognition specificity.

Analysis of the Binding Properties of Cre Mutants in Vitro—Fig. 1 depicts the results of a band shift assay of symmetrical loxP sites with symmetrical single base substitutions in position 2 or position 6 of the inverted repeats and Cre mutants. The comparison of lanes 6 or 7, lanes 22 and 23, respectively (loxP variants and Cre wild-type) to lane 5 or lane 21 (loxP and Cre wild-type) reveals the complete loss of DNA binding of Cre upon changes in positions 2 and 6 of loxP. Complexes I and II were denoted corresponding to Mack et al. (13). The band corresponding in our analysis to complex I (one Cre molecule bound) was identified by using a loxP half-site (lanes 4 and 20).

Fig. 3 summarizes the results of gel retardation experiments with various Cre mutants. In this table only the results of the analysis with positions 2 and 6 of the inverted repeats of loxP are included, since these positions have been shown to be the most sensitive ones with respect to the DNA recognition of Cre (Fig. 2). Positions outward from 7 are of minor importance for the attachment of Cre to its target and have been omitted from this analysis (Fig. 2).

In general mutations in helix K of Cre (residues 294–303) showed only very minor differences in their recognition pattern from the wild-type protein with respect to the binding specificity (Fig. 3A). In most cases the binding affinity of the respective mutant to wild-type target is reduced. As an example mutant R297T is shown in Fig. 1 (lanes 8–10). Substitutions in helix M, in the vicinity of the conserved Tyr-324 had a more significant effect (Fig. 3B). Substitutions of Asn-317 to Ala and substitutions of Asn-319 to Cys, Gly, His, Leu, Pro, Gln, Ser, and Thr resulted in a broadening of binding specificity at target positions 2 and/or 6. Three mutants, N317A, N319C, and N319P, are shown as examples for broadening of binding specificity in Fig. 1 (lanes 11–13, 24–26, and 30–32, respectively).

In contrast to these results, mutations in position 318 resulted in a reduced binding affinity of complex II but not of complex I to wild-type target (Fig. 1, lanes 14–16). Binding to target variants in positions 2 and 6 is abolished (Fig. 3B).

Comparison of the Binding Properties of Cre Mutants in Vitro and the Recombination Events in Vivo—Fig. 4 summarizes the results of the band shift assays in comparison to the results of the in vivo recombination test. The in vitro test on minimal lactose plates relies on the selection of those cells where Cre-mediated recombination has removed two stop codons, flanked by two loxP sites, within the coding region of lacZ located on a reporter plasmid (Fig. 1). These stop codons inhibited expression of β-galactosidase and Lac permease and thereby cell growth on minimal lactose plates.

The results of the in vitro test, the band shift assays, can be compared with the in vivo results only with caution. First, the in vitro gel retardation experiments in contrast to the in vivo recombination experiments were done in the presence of heparin as competitor. But the differences between the two test systems go beyond the difference between target recognition by the protein and the sequence of events characterizing the recombination process. With the in vitro test, binding of Cre to the respective target sequence is determined. With respect to Cre, this reaction is an equilibrium between two states of the protein, DNA bound and non-bound. In contrast to a balanced equilibrium, the in vivo system selects for a single recombination event in a single cell. This single event will not be reversed and results in the survival of only the respective cell under the given conditions. Fig. 4 reflects the differences in the two test systems.

Comparison of the target recognition and recombination (Fig. 4) of Cre wild-type to the mutant R292L reveals the difference between “simple” DNA binding and full recombination. Both proteins bind the wild-type loxP target nearly equally well (Fig. 4, indicated by the number of plus signs). However, only wild-type Cre mediates recombination of the respective target site. With R292L recombination is completely abolished (Fig. 4: confluent growth of colonies versus no colonies). The substitution of Arg-292 by leucine has very minor effect on the recognition and binding of the target sequence, which precedes the recombination event. But as Arg-292 is a member of the catalytic quartet, the recombination event itself is inhibited by the replacement.

The difference between the single, non-reversible event and an equilibrium can be illustrated with N319L. This mutant mediates recombination with loxP2 (24 recombinant colonies per plate), although the protein does not seem to bind this target in vitro (Fig. 4). Nevertheless, mutations in the more N-terminally located helix K (mutations at positions 292–301) in general show no or reduced affinity to the targets we tested with respect to wild-type Cre. This reduced affinity in general is reflected in the in vivo test: only a few colonies are formed on selection plates. In contrast to these mutants, amino acid substitutions in the region surrounding helix M (substitution at positions 317–323) reveal a broadening or shift of recombina-
**Fig. 3.** Retardation of symmetrical \(\text{loxP}\) sites with single base changes in positions 2 and 6 in the inverted repeats by Cre mutants.

The sequence of only one of the inverted repeats of \(\text{loxP}\) is shown. Numbers of the bases are the same as in Fig. 2. The relative amount of shifted

| Pos. | w1 | 289 | 290 | 291 | 292 | 293 | 294 | 295 | 296 | 297 | 298 | 299 | 300 | 301 | 302 |
|------|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|      |    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| R299L | H  | S   | A   | R   | V   | G   | A   | R   | D   | M   | A   | R   | A   |     |     |
| R297L |    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| R297T | L  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| R295L | T  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| V293H, R297T | L |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| V293R, R297T | R  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

| Pos. | w1 | 317 | 318 | 319 | 320 | 321 | 322 | 323 | 324 | 325 | 326 | 327 |
|------|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|      |    |     |     |     |     |     |     |     |     |     |     |     |
| N317A | A  |     |     |     |     |     |     |     |     |     |     |     |
| N317L | L  |     |     |     |     |     |     |     |     |     |     |     |
| N317R | R  |     |     |     |     |     |     |     |     |     |     |     |
| V318K | K  |     |     |     |     |     |     |     |     |     |     |     |
| V318R | R  |     |     |     |     |     |     |     |     |     |     |     |
| N319A | A  |     |     |     |     |     |     |     |     |     |     |     |
| N319C | C  |     |     |     |     |     |     |     |     |     |     |     |
| N319D | D  |     |     |     |     |     |     |     |     |     |     |     |
| N319E | E  |     |     |     |     |     |     |     |     |     |     |     |
| N319F | F  |     |     |     |     |     |     |     |     |     |     |     |
| N319G | G  |     |     |     |     |     |     |     |     |     |     |     |
| N319H | H  |     |     |     |     |     |     |     |     |     |     |     |
| N319I | I  |     |     |     |     |     |     |     |     |     |     |     |
| N319K | K  |     |     |     |     |     |     |     |     |     |     |     |
| N319L | L  |     |     |     |     |     |     |     |     |     |     |     |
| N319M | M  |     |     |     |     |     |     |     |     |     |     |     |
| N319P | P  |     |     |     |     |     |     |     |     |     |     |     |
| N319Q | Q  |     |     |     |     |     |     |     |     |     |     |     |
| N319R | R  |     |     |     |     |     |     |     |     |     |     |     |
| N319S | S  |     |     |     |     |     |     |     |     |     |     |     |
| N319T | T  |     |     |     |     |     |     |     |     |     |     |     |
| N319U | U  |     |     |     |     |     |     |     |     |     |     |     |
| N319V | V  |     |     |     |     |     |     |     |     |     |     |     |
| N319W | W  |     |     |     |     |     |     |     |     |     |     |     |
| N319Y | Y  |     |     |     |     |     |     |     |     |     |     |     |
| M322R | R  |     |     |     |     |     |     |     |     |     |     |     |
| M322K | K  |     |     |     |     |     |     |     |     |     |     |     |
| N322K | K  |     |     |     |     |     |     |     |     |     |     |     |
| N322L | L  |     |     |     |     |     |     |     |     |     |     |     |
| N322R | R  |     |     |     |     |     |     |     |     |     |     |     |
| N322L | L  |     |     |     |     |     |     |     |     |     |     |     |
| N322L | L  |     |     |     |     |     |     |     |     |     |     |     |

The sequence of only one of the inverted repeats of \(\text{loxP}\) is shown. Numbers of the bases are the same as in Fig. 2. The relative amount of shifted
Cre Mutants with Altered DNA Binding Properties

![Image of Cre Mutants](image)

**DISCUSSION**

The analysis of the crystal structure of the Cre-lox complex revealed abundant data concerning possible protein-DNA interactions (18). Nevertheless, it must be noted that these DNA contacts refer to a protein monomer that has already cut the underlying DNA strand (18). Guo et al. (18) described an intramolecular rearrangement of Cre after its initial binding to the target sequence. As a consequence of this rearrangement, the initial recognition and initial binding specificity might be not exclusively determined by the amino acids listed by Guo et al. (18).

Cre consists of two domains, one N-terminal domain from amino acid 20–129 and a C-terminal domain from amino acid 132–341. Both together form a clamp around the target DNA (18). To analyze the target recognition specificity of Cre, we constructed mutants of this protein with amino acid substitutions in different positions of the presumptive DNA binding region in the C-terminal part of Cre. Retardation assays with wild-type Cre revealed that base pairs 2 and 6 of the inverted repeats of loxP are the most sensitive positions with respect to protein recognition (Fig. 2). This agrees well with the crystal structure as follows: two mostly helical regions in the N-terminal domain of Cre, amino acid positions 37–50 and positions 81–101 seem to interact with base positions 1 and/or 2. Bases 5 and/or 6 seem to interact with positions 257–262, another stretch in which its major adopts a-helical conformation and is located in the C-terminal domain of Cre. Nevertheless, the contacts from the N-terminal domain of Cre mentioned cannot be the only determinants for the sensitivity of Cre with respect to base pair 2, because hydroxyl radical footprints with a C-terminal fragment of Cre, the so-called 25-kDa fragment (amino acids 118 and/or 119–343), revealed, in comparison to footprint prints with the full size wild-type protein, only differences in protection within the 8-bp spacer region of loxP (11). No differences outside the spacer region were observed, particularly no differences in position 2. Therefore, the recognition of base position 2 cannot be restricted exclusively to amino acids in the N-terminal domain. Since the 25-kDa fragment approximately corresponds to the C-terminal domain of Cre, the results from the footprint experiments (11) agree well with the crystal data, where the spacer region is not contacted by the C-terminal domain. From the crystal structure and as concluded from our biochemical data, we suggest that base pair 2 of the inverted repeats, in addition to amino acids of the N-terminal domain of Cre, is specifically bound by amino acids of helix M and/or amino acids of the loop connecting helices L and M. The recognition specificity may switch to only the N-terminal amino acids after the DNA strand has been cut and linked to Tyr-324.

Amino acid substitutions in helix M and its N-terminal loop (positions 317–323) often result in a broadening of the recognition specificity (Fig. 4). However, we did not observe a complete switch of the target recognition specificity with any of the mutants we tested.

A shift of recognition specificity is, of the mutants we tested, best illustrated by N317A (Fig. 4). This mutant is characterized by a strong reduction in recombination efficiency with wild-type loxP in comparison to wild-type Cre. But, in contrast to wild-type Cre, which does not mediate recombination with target DNA in each case was determined from at least four independent experiments: +++, 76–100% retardation as compared with the wild-type complex; +++, 56–75% retardation as compared with the wild-type complex; +, 26–55% retardation as compared with the wild-type complex; +, 5–25% retardation as compared with the wild-type complex; –, less than 5% retardation as compared with the wild-type complex. All retardation values are given with respect to wild-type complex II, with two molecules of Cre bound to one loxP site. A, amino acid substitutions in helix K. The amino acid sequence of Cre wild-type from position 317 to 327 (helix M) is given, with the amino acid replacements in each mutant indicated below: B, amino acid substitutions in helix M. The amino acid sequence of Cre wild-type from position 317 to position 327 (helix M) is given, with the amino acid replacements in each mutant indicated below.
bination specificity (Fig. 4). This broadening is most obvious with the mutants N317A, N319L, N323R, and target variants in position 6 or position 2 of the inverted repeats. These results might be due to direct amino acid-base pair contacts. Nevertheless, it has to be kept in mind that, especially with N323R, an amino acid adjacent to a residue of catalytic function, Tyr-324, has been modified. The broadening in recombination specificity might therefore be the result of indirect effects of the amino acid replacement on structure and orientation of the catalytic center. This explanation is supported by the in vitro results with substitutions in amino acid position 318; both mutants, V318R (Fig. 1, lane 14) and V318K (data not shown), resulted in vitro in only the formation of complex I. We could not detect bands corresponding to complex II in the retardation assay (Fig. 1, lane 14). With both substitutions, the in vitro recombination efficiency, even with wild-type loxP, is strongly reduced (Fig. 4). Both mutations in position 318 that we have tested might modify the spatial orientation of helix M toward the DNA, thereby inhibiting the binding of a second Cre monomer to the loxP site.

Surprisingly, in contrast to position 318, substitutions in position 319 do not inhibit the formation of complex II (Fig. 1, lanes 24, 27, and 30). Depending on the amino acid that has been inserted instead of the wild-type asparagine, in most cases we observed broadening of recognition specificity at position 2 in vitro. The replacement of Asn-319 by Pro, Gln, or Gly in vitro affected also position 6 of the inverted repeats (Fig. 3B). In vitro all mutants we tested in position 319 recombined with the target variant in position 2. Only N319L and N319P used the target variant in position 6 for recombination. Nevertheless, with the exception of N319L, the recombination efficiency of the mutants with the target variants was rather low.

The additive, clamp-like interaction of both domains of Cre with the target sequence seems to be necessary to sufficiently stabilize the recombination complex and ensure the correct orientation of both DNA strands during strand exchange. A consequence of this binding mode is the prediction that in order to design a Cre recombinase with a complete change in DNA binding specificity, at least two amino acids, one in the N-terminal and one in the C-terminal domain of the protein, will have to be replaced. According to Fig. 2, those parts of the protein that contact bases outward from position 7, presumably positions 240–250, can be neglected in these experiments. The DNA contacts that are expected in this region (sheets 4 and 5, Guo et al. (18)) should be nonspecific contacts, e.g. to the DNA backbone.

Since the structure of the Cre-loxP complex was published only recently (18), we planned and carried out all amino acid replacements based on the alignments of the amino acid sequences and of the secondary structures of other integrases to the predicted secondary structure of Cre (λ Int (20), HP1 integrase (19)). Since none of these “reference integrases” had been crystallized as a co-complex with its DNA target, we had to deduce the amino acid positions that might be involved in DNA recognition from the x-ray structure of γδ resolvase, which was crystallized bound to its target sequence (24). After the publication of the x-ray structures of Cre (18) and XerD (21), both proteins were included in the alignment of the secondary structures and confirmed the match between x-ray structure and structure prediction. For each protein, we checked to what extent those secondary structure elements, which had been predicted, matched with those secondary structure elements that were deduced from the x-ray structures.

The C-terminal region surrounding the catalytic triad in all integrases seems to adopt a conserved three-dimensional structure. This structure is even comparable to the structure of γδ resolvase. Besides huge differences between the resolvases/invertases and the integrases, at least with respect to the location of the catalytic site, γδ resolvase seems to adopt a triple helical motif similar to the integrases for specific recognition of its target sequence (24).

In contrast to all other integrases of known structure and in contrast to the secondary structure prediction, a flexible loop around the catalytic tyrosine 342 is formed in λ integrase. The remaining part of the catalytic core of λ integrase adopts the same secondary structure as is found in the other integrases. Since λ integrase had been crystallized without its target DNA, we predict, in analogy to the rearrangement of bZIP proteins (42, 43), a structural rearrangement or induced fit (44) of the flexible loop of λ integrase in the presence of its DNA target. The flexible loop of λ integrase should reorganize toward an α-helix, which sterically adjusts the catalytic triad to the conformation found with HP1, XerD or, Cre. This should result in a triple helical DNA binding motif which is common for all integrases and underlines recent results with eukaryotic topoisomerases and site-specific recombinases that suggest that the catalytic domains of these proteins derive from a common ancestral strand transferase (45).

Acknowledgments—We thank Benno Müller-Hill and Andrew Barker for stimulating discussions and critically reading this manuscript, and Karin Otto and Gudrun Zimmer for excellent technical assistance with the synthesis of oligonucleotides.

REFERENCES

1. Stark, W. M., Boocock, M. R., and Sherratt, D. J. (1992) Trends Genet. 8, 422–429
2. Grindley, N. D. F. (1997) Curr. Biol. 7, R608–R612
3. Argos, P., Landy, A., Abremski, K., Egan, J. B., Ljungquist, E. H., Hoess, R. H., Kahn, M. L., Kalonis, B., Narayana, S. V. L., Pierson, L. S., Sternberg, N., and Leong, J. M. (1986) EMBO J. 5, 433–440
4. Abremski, K., and Hoess, R. (1992) Protein Eng. 5, 87–91
5. Sternberg, N. (1978) Cold Spring Harbor Symp. Quant. Biol. 43, 1143–1146
6. Sternberg, N., Austin, S., Hamilton, D., and Yarmolinsky, M. (1978) Proc. Natl. Acad. Sci. U. S. A. 75, 5594–5598
7. Sternberg, N., and Hamilton, D. (1981) J. Mol. Biol. 150, 467–487
8. Sternberg, N., Sauer, B., Hoess, R., and Abremski, K. (1986) J. Mol. Biol. 187, 197–212
9. Abremski, K., and Hoess, R. (1984) J. Biol. Chem. 259, 1509–1514
10. Abremski, K., Hoess, R., and Sternberg, N. (1983) Cell 32, 1301–1311
11. Hoess, R., Abremski, K., Irwin, S., Kendall, M., and Mack, A. (1990) J. Mol. Biol. 216, 873–882
12. Hoess, R., Ziese, M., and Sternberg, N. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 3398–3402
13. Mack, A., Sauer, B., Abremski, K., and Hoess, R. (1992) Nucleic Acids Res. 20, 4451–4455
14. Hoess, R. and Abremski, K. (1985) J. Mol. Biol. 181, 351–362
15. Hoess, R., Wierzchicki, A., and Abremski, K. (1986) Nucleic Acids Res. 14, 2287–2300
16. Sauer, B. (1987) Mol. Cell. Biol. 7, 2087–2096
17. Gu, H., Marsh, J. D., Orban, P. C., Bossmann, H., and Rajewsky, K. (1994) Science 265, 103–106
18. Guo, F., Gopolal, D. N., and van Duyne, G. D. (1997) Nature 389, 40–46
19. Hickman, A. B., Waninger, S., Scocca, J. J., and Dyda, F. (1997) Cell 89, 257–267
20. Kwon, H. J., Tirumalai, R., Landy, A., and Ellenberger, T. (1997) Science 276, 126–131
21. Subramanuy, H. S., Ariceszewska, L. K., Baker, R. A., Bird, L. E., Sherratt, D. J., and Wigley, D. B. (1997) EMBO J. 16, 5178–5187
22. Hoess, R., Wierzchicki, A., and Abremski, K. (1989) in Structure and Methods: Human Genome Initiative and DNA Recombination (Sarma, R., and Sarma, M., eds) Vol. 1, pp. 203–213, Adenine Press, Schenectady, NY
23. Wierzchicki, A., Kendall, M., Abremski, K., and Hoess, R. (1987) J. Mol. Biol. 195, 785–784
24. Yang, W., and Steitz, T. A. (1995) Cell 82, 193–207
25. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) Methods Enzymol. 185, 60–89
26. Oehler, S., Eismann, E. R., Kramer, H., and Muller-Hill, B. (1990) EMBO J. 9, 937–939
27. Hanahan, D. (1985) DNA Cloning: A Practical Approach, Vol. 1, pp. 109–114, IRL Press at Oxford University Press, Oxford
28. Fellenberg, K. (1998) Lokale Gonaktivierung in der Haut. M.Sc. thesis, Institute for Genetics, University of Cologne
29. Chang, A. C., and Cohen, S. N. (1978) J. Bacteriol. 134, 1141–1156
30. Rose, R. E. (1986) Nucleic Acids Res. 16, 355
Cre Mutants with Altered DNA Binding Properties

31. Oehler, S., Amouyal, M., Kolkhof, P., von Wilcken-Bergmann, B., and Müller-Hill, B. (1994) *EMBO J.* **13**, 3348–3355
32. Kolkhof, P., and Müller-Hill, B. (1994) *J. Mol. Biol.* **242**, 23–26
33. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., A.1–A.3, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
34. Miller, J. H. (1972) *Experiments in Molecular Genetics*, pp. 432–433, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
35. Laemmli, U. K. (1970) *Nature* **227**, 680–685
36. Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248–254
37. Hanahan, D. (1983) *J. Mol. Biol.* **166**, 575–580
38. Rost, B., and Sander, C. (1993) *J. Mol. Biol.* **232**, 584–599
39. Rost, B., and Sander, C. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 7558–7562
40. Rost, B., and Sander, C. (1994) *Proteins* **19**, 55–72
41. Rost, B., Sander, C., and Schneider, R. (1994) *Comput. Appl. Biosci.* **10**, 53–60
42. O’Neil, K. T., Huess, R. H., and DeGrado, W. F. (1990) *Science* **249**, 774–778
43. O’Neil, K. T., Shuman, J. D., Ampe, C., and DeGrado, W. F. (1991) *Biochemistry* **30**, 9030–9034
44. Spolar, R. S., and Record, M. T., Jr. (1994) *Science* **263**, 777–784
45. Cheng, C., Kussie, P., Pavletich, N., and Shuman, S. (1998) *Cell* **92**, 841–850