UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 582

Endonuclease II - a GIY-YIG enzyme of bacteriophage T4

PERNILLA LAGERBÄCK

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2008

ISSN 1651-6214
ISBN 978-91-554-7361-7
urn:nbn:se:uu:diva-9410

Dissertation presented at Uppsala University to be publicly examined in B41, Biomedicinskt Centrum, Husargatan 3, Uppsala, Friday, December 19, 2008 at 10:15 for the degree of Doctor of Philosophy. The examination will be conducted in English.

Abstract
Lagerbäck, P. 2008. Endonuclease II - a GIY-YIG enzyme of bacteriophage T4. Acta Universitatis Upsaliensis. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 582. 53 pp. Uppsala. ISBN 978-91-554-7361-7.

Endonuclease II (EndoII) of bacteriophage T4 is a GIY-YIG enzyme involved in host DNA breakdown during phage infection of *E. coli*. EndoII combines features of restriction endonucleases with those of homing endonucleases in that it breaks down DNA foreign to itself but recognizes a 16 bp long asymmetric and ambiguous sequence. This investigation addresses the biological function of EndoII, its mode of interaction with its substrate and roles of individual residues in catalysis, sequence recognition and binding.

It is shown here that EndoII increases the frequency of non-homologous recombination in phage-infected cells, showing that EndoII indeed can induce recombinational events. Although single-stranded nicks are frequent in *in vitro* reactions with purified protein, the enzyme is found to produce mostly double-stranded breaks *in vivo*, since nicks are repaired. Mutations of residues positioned on the putative catalytic surface result in severely reduced catalytic activity, while residues in the N-terminal region and a middle region (MR) appear to mainly contribute to substrate binding. Mutation of the putatively magnesium-binding residue E118 renders the enzyme catalytically inactive. Residues K76 (in the MR and positioned on the catalytic surface) and G49 and R57 (on the catalytic surface) also contribute to substrate recognition. All mutants bind as tetramers to two DNA molecules, indicating that the wildtype would also bind as a tetramer. EndoII E118A alone can bind also in monomeric and dimeric form to one DNA molecule, possibly because the glutamate charge normally repels the DNA. The solved crystal structure of tetrameric EndoII E118A shows a striking X-shape with two putative catalytic surfaces to each side positioned so that double-stranded cleavage would require severe DNA distortion. Combination of all data suggests that upon binding *in vivo* EndoII scans the DNA for a second binding site, binding to both sites but nicking or cleaving only one of them.

Keywords: GIY-YIG, EndoII, endonuclease, structure, tetramer, binding, nicking, recombination

Pernilla Lagerbäck, Department of Cell and Molecular Biology, Box 596, Uppsala University, SE-75124 Uppsala, Sweden

ISSN 1651-6214
ISBN 978-91-554-7361-7
urn:nbn:se:uu:diva-9410 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-9410)

”Det är klokt att veta vad man letar efter innan man börjar leta.”

Nalle Puh

To my family

List of Publications

This thesis is based on the following papers, which will be referred to in the text by their Roman numerals:

I Bacteriophage T4 endonuclease II: concerted single-strand nicks yield double-strand cleavage.
Carlson, K, **Lagerbäck, P** and Nyström, AC
Molecular Microbiology 52: 1403-11 (2004)

II Amino acid residues in the GIY-YIG endonuclease II of phage T4 affecting sequence recognition and binding as well as catalysis.
Lagerbäck, P and Carlson K
Journal of Bacteriology 190: 5533-44 (2008)

III Bacteriophage T4 endonuclease II, a promiscuous GIY-YIG endonuclease, binds as a tetramer to two DNA substrates.
Lagerbäck, P, Andersson, CE and Carlson, K
Manuscript 2008

IV Bacteriophage T4 endonuclease II enhances recombination between phage and plasmid DNA.
Carlson, K and **Lagerbäck, P**
Manuscript 2008

V Structure of bacteriophage T4 endonuclease II mutant E118A, a tetrameric GIY-YIG enzyme.
Andersson, CE, **Lagerbäck, P** and Carlson, K
Manuscript 2008

Reprints were made with the permissions from the publishers.

Contents

Abbreviations

AdoMet	S-adenosylmethionine
ATP	adenosine triphosphate
bp	base pair
CTP	cytosine triphosphate
DNA	deoxyribonucleic acid
EndoII	endonuclease II
gp	gene product
GTP	guanosine triphosphate
HE	homing endonuclease
kDa	kilodalton
MR	middle region
NTR	N-terminal region
ORF	open reading frame
RE	restriction endonuclease

Introduction

Protein-DNA interaction is a central phenomenon in any living cell. These interactions are involved in fundamental events like replication, transcription, gene regulation, defense against foreign DNA and more. By understanding how these proteins work we are one step closer to understanding a fundamental part of life. Much is already known about the numerous ways in which these interactions take place, but even more remains to be discovered. Endonucleases are enzymes that cleave DNA and can be part of DNA repair systems, systems for propagating the gene encoding the endonuclease, DNA replication, defense systems and more.

In this thesis Endonuclease II (EndoII) from bacteriophage T4 has been closely studied regarding its interactions with the DNA sequence that it recognizes, as well as its biological role.

Bacteriophage T4

Bacteriophages (from "bacteria" and the greek word "phagein", to eat) are viruses that infect bacteria. They were first identified by Frederick Twort in 1915, but were independently identified, properly characterized and named by Félix d'Herelle (1873-1949) in 1917 (180). d'Herelle isolated phages from stool samples of patients recovering from dysentery and showed that these unknown microbes could lyse cultures of Shiga bacilli and that they also grew in number in these cultures. Since their discovery phages have been widely used as research tools in molecular biology for studying for example replication, transcription and regulation. After a period of relative uninterest phages are again in the hot spot for research, among other things as a substitute for antibiotics.

Bacteriophage T4 (Fig. 1) belongs to the large tailed coliphages of the T-even group in the family Myoviridiae. Studies of T4 have been performed since the 1930's and it is one of the best characterized phages. The T4 genome has been completely sequenced (GenBank AF158101), as have the genomes of several T4-like phages (http://phage.bioc.tulane.edu/). Its capsid is composed of an icosahedral head that contains the large, linear double-stranded genome, and a tail with tail fibers that determine the host specificity. The tail consists of a tube surrounded by a contractile sheath, the genome

is released into the bacterium through this tube after it has penetrated the bacterial envelope (156).

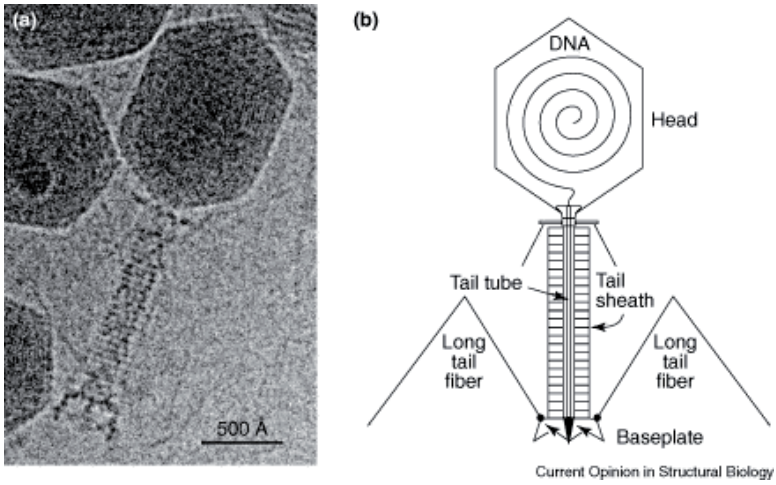

Figure 1. (a) Cryo-EM micrograph of bacteriophage T4. (b) Schematic picture of the major structural components. The icosahedral head contains the large double-stranded genome which is delivered into the host through the internal tail tube that is inside the contractile sheath. The bacterial envelope is recognized by the extended tail fibers. The short tail fibers are shown as bent arrow-like objects around the periphery of the baseplate. The black triangle in the middle of the baseplate represents the cell-puncturing device. Reprinted from Rossman *et al.* (156) with permission from the publisher and Paul Chipman who provided the cryo-EM micrograph.

The T4 genome

The T4 genome consists of 168,903 bp (GenBank AF158101), predicted to contain approximately 300 genes of which 289 are probable protein encoding genes, 8 are tRNA genes and 2 encode small stable RNAs of unknown function (117). Of these 300 predicted genes approximately 156 had been characterized in 2003, while the rest - almost half of the T4 genes - had not been assigned a function. The T4 genome has a very high gene density and the arrangement is very complex, with proteins encoded from different start points in the same gene; genes containing introns; genes within genes but with different reading frames, on the same or on opposite strands; translational bypassing, where part of the mRNA in a coding region is not translated. The high gene density is also the result of the scarcity of regions in the T4 genome that are non-coding, only 9 kb or 5.3 % of the entire genome. It is not unusual that a termination codon overlaps with the start codon of the following gene. Also the regulatory regions are compact, sometimes overlapping with coding regions. T4 genes are also very small in general, 62 of the predicted proteins contain less than 100 amino acids.

By using hydroxymethylated cytosines instead of cytosines, phage DNA is protected from degradation by T4 nucleases and host encoded nucleases that attack the phage DNA during infection. The cytosine bases are hydroxymethylated by T4 enzymes prior to their incorporation into the genome (200, 201). This is quite unusual since in other organisms most modifications of DNA are performed after incorporation. These hydroxymethylated cytosines are further modified by glucosylation by α- or β-glucosylase (70% and 30% respectively) (92, 93). Due to the *alc* gene product hydroxymethylation is also essential for the transcription of middle mode genes, since the product of this gene inhibits transcription of cytosine DNA (85, 171, 172).

T4 infection

Bacteriophage T4 infects *Escherichia coli* and closely related *Shigella* species, but also some isolates of the plant pathogen *Erwinia* (146). There are three distinct phases of infection: at 37˚C the immediate early phase begins within 30 seconds after infection, the middle phase after 3 to 4 minutes (before replication) and the late phase after DNA replication is initiated (123). The phage particle recognizes its target through the long tail fibers that attach to receptors (lipopolysaccharide (LPS) or OmpC) on the surface of the bacteria (66, 202). When at least three of these long fibers have bound (39), the short tail fibers are extended and bind irreversibly to LPS (150). The baseplate goes through a conformational change to drive the tail tube through the outer membrane. Lysozyme from the tip of the tail is activated and breaks down the peptidoglycan layer to create an opening to the cytoplasmic membrane (40, 83). The genome is then released and transferred into the cytoplasm of the bacterium.

Early in infection only early T4 promoters are transcribed. DNA, RNA and protein synthesis of the host cell is shut off almost immediately, and already after 1 minute (at 37˚C) there is almost no new synthesis of host proteins (37). The transition from host to phage protein synthesis is performed by a number of different T4 enzymes. For example, the **Alc** (*a*borts e*l*ongation on *C*yt-DNA) protein terminates the elongation of transcription of cytosine-containing DNA (i.e. host DNA) (47, 97-99). The ADP-ribosyltransferase **Alt**, which is packaged in the phage particle and injected with the phage DNA, modifies the RNA polymerase of the host which leads to a stronger interaction with the T4 **early** promoters (88, 173). The ***ndd*** (*n*uclear *d*isruption *d*eficient) gene product is responsible for the disruption of the host nucleoid that delocalizes the host DNA from the middle of the cell to the bacterial envelope, it also inhibits host DNA replication by blocking the replication forks (19, 20). **EndoII** is also part of the host shutoff, degrading the host DNA together with **EndoIV** and the ***46/47*** exonuclease (see below) (157). There are several additional T4 proteins that contribute to

host shutoff or modification among the early expressed proteins that have not yet been characterized (117).

Phage DNA replication starts 4-6 minutes after infection, using a number of phage-encoded proteins like e.g. T4 DNA polymerase (product of gene *43*, gp43), single-strand DNA binding protein (gp32), endonuclease VII (gp49), the 46/47 exonuclease and DNA ligase (gp30) (95, 131). Replication rounds can be initiated in two ways. The early stage of replication is origin-dependent, initated at several specific origins; this also involves RNA polymerase. At later times recombination-dependent replication is dominating. Recombination-dependent replication enables replication of the 3' ends of the linear genome by formation of concatemers through recombination between chromosome ends and homologous sequences in other phage DNA molecules. These concatemers can be up to 20 genomes long. This action involves the 46/47 exonuclease and can be initiated with or without EndoVII (13, 23, 72, 95). Hydroxymethylation of cytosines, before incorporation into the phage DNA, requires dCTPase (gp56), dCMP hydroxymethylase (gp42) and deoxynucleotide kinase (gp1) for *de novo* synthesis. In the absense of dCTPase hydroxymethylated cytosines can be produced from the nucleotides released from the host DNA using only dCMP hydroxymethylase and deoxynucleotide kinase. If hydroxymethylation is disturbed it will result in phage carrying cytosine DNA; however, this also requires inactive Alc to allow late transcription of cytosine DNA (122) and inactive endonuclease II to prevent DNA degradation (47).

The ADP-ribosyltransferase **ModA**, like Alt, ribosylates the host RNA polymerase. The action of ModA, however, leads to inhibited transcription of T4 early promoters and transcription is instead concentrated to **middle** mode promoters (60, 186, 187). The ribosylation reduces the affinity of the host RNA polymerase for the host σ^{70} factor, and **AsiA** also helps in reducing its affinity for -35 promoter elements by binding to it (134, 166). Middle mode transcription also requires **MotA** which acts like a bridge between middle promoter -30 elements (MotA box) and σ^{70}, which activates transcription from middle mode promoters (17, 135).

Transcription of **late** promoters results in the production of phage capsid particles and proteins required for virion assembly. The phage encoded sigma factor (gp55) associates with the RNA polymerase and directs it to late promoters (86), while gp33 is necessary for enhancing replication by promoting the interaction of the RNA polymerase with the translocating DNA replication fork (68, 69). To assemble mature virions the DNA is packaged into the icosahedral heads, and when the head is full the DNA is cut off and the last components of the head are added. The length of the packaged DNA is approximately 102% of a genome-equivalent length, resulting in a genome with terminal redundancy and endpoints that vary between the phage particles. Then the assembled tails are attached and finally the six tail fibers. As the last stage of infection (15-30 minutes after infection

at 37˚C) the bacterium is lysed. The T4 holin (gpt) (109) creates a hole in the inner membrane, allowing access for the T4 lysozyme (gpe) (127) to break down the peptidoglycan layer. However, if the cell culture is dense, lysis can be delayed for several hours ("lysis inhibition") (2, 31). After lysis has occured hundreds of new phage are released, ready to infect the next bacterium and produce another generation of phage.

Host DNA degradation in T4 infected cells

Fragmentation of T4 cytosine-DNA and bacterial DNA *in vivo* is visible within 4-10 minutes after infection (33, 100). This breakdown of host DNA is initiated by EndoII (67, 158, 159, 194). Crude extracts withdrawn at different timepoints after T4 infection exhibited maximal EndoII activity on λ DNA 10 minutes after infection. EndoII activity decreased again approximately 15 minutes after infection (157). The large (approximately 100-1000 bp) DNA fragments produced by EndoII are then degraded exonucleolytically by the 46/47 exonuclease (also active in recombination-dependent replication). Single-stranded DNA is attacked by endonuclease IV (EndoIV) which cleaves single-stranded DNA 5' of cytosines at the preferred sequence TCA (71, 132, 158). EndoII, although dispensable for T4 infection, is important for efficient degradation of host DNA while the EndoIV protein is involved to a minor extent (32, 33). EndoIV has been suggested to have its major function as an inhibitor in the replication of host DNA (32, 71). The 46/47 exonuclease however, is essential for producing free nucleotides that can be incorporated into the phage genome (100, 198). The degradation of host DNA completely abolishes the possibility for the host to produce new enzymes that might strike back at the phage, and nucleotides from degraded host DNA can produce 20-30 new phage (67), however, in rich media the majority of nucleotides for phage replication come from *de novo* synthesis.

Sequence-specific endodeoxyribonucleases

Sequence-specific endonucleases, like e.g. EndoII and EndoIV, can have various roles despite their seemingly similar action - DNA cutting. The substrate recognition, the cutting pattern and the physiological consequences of endonuclease action can be and are widely diverse. To find their cognate DNA sequences proteins can either scan the DNA by linear diffusion (1D diffusion) until a sequence is reached to which the protein can bind specifically, or by dissociating and diffuse through space to reassociate through random collision (3D diffusion), most likely on the same DNA molecule (12). A third way, called intersegmental transfer, occurs when a protein capable of binding to two sites binds transiently to the second site before dissociating from the first. When a specific site is found, the protein binds to

the DNA either contacting the base edges of specific bases (direct recognition) or by contacting the DNA backbone (indirect recognition). The most common contacts are made through hydrogen bonds (50%), most often involving hydrophilic amino acids, especially the positively charged amino acids lysine and arginine. These residues most often contact phosphate groups of the DNA backbone (which in general is the most common interaction between protein and DNA), but also base edges (104). Other interactions in decreasing order of occurrence are van der Waals, hydrophobic, and electrostatic interactions (104). In addition, water mediated contacts are very important and can have varying roles (79).

Another important part for recognition is the ability of the DNA to be distorted upon specific binding, e.g. HincII distorts the DNA by bending, unwinding and shifting of base planes (6). This is influenced by the base sequence and is almost always the result of positive roll at pyrimidine-purine base steps (45). Also the protein is subjected to conformational changes, for example NotI changes conformation upon binding, inserting a large number of residues into the major groove to make several base-specific contacts (102), and HincII has a more closed structure when bound to DNA (106).

Homing endonucleases

Nucleases can be grouped together either based on shared amino acid sequence motifs, and thus structural motifs, or based on their biological function. Homing endonucleases (HEs) are enzymes that promote the mobility of the DNA sequence that encodes them and sequences close to these. They are often located within an intron or intein and can promote the homing event of the entire intron (e.g. like the *td* intron encoded I-TevI (36, 197)). They also have been found as freestanding ORFs promoting the mobility of the ORF and surrounding sequences, e.g. SegG (107) and SegF (11). HEs nick or cleave within or outside a recognition sequence (typically 14-40 bp long, allowing for some sequence variation) in alleles that do not contain a copy of the HE coding sequence. The cleavage sequence is usually located close to the intron insertion site as is the case for e.g. I-CeuI and I-CreI (3, 51, 113, 114, 185, 189), but can also be several base pairs away from this as for I-TevIII (52). The break introduced by the HE initiates DNA repair pathways that will seal the break by homologous recombination, thus introducing the HE gene also into the new allele (group I introns and inteins). A more complex pathway involving DNA cleavage, reverse splicing and reverse transcription of a DNA-RNA hybrid also exists (group II introns) (178).

Homing endonucleases (HEs) can be divided into five groups depending on their conserved sequence elements: LAGLIDADG, HNH, HisCys box, GIY-YIG or PD-(D/E)-XK (178).

LAGLIDADG

LAGLIDADG endonucleases are most often found within introns or inteins in mitochondrial or chloroplast genomes of single cell eukaryotes (178). They appear either as homodimers containing one LAGLIDADG motif in each subunit, or as monomers containing two LAGLIDADG motifs, to be able to cleave both strands of the DNA (for examples see I-CreI (193) and I-DmoI (1) respectively). Each LAGLIDADG motif constitutes one active site that cleaves one strand of the recognition sequence, two LAGLIDADG motifs are thus required for double-stranded clavage. The recognition sequences of LAGLIDADG homodimeric enzymes, like I-CreI (51, 185, 193) and I-CeuI (114) are palindromic or near-palindromic, while those of monomeric LAGLIDADG enzymes, like I-DmoI (1) do not require symmetric DNA targets. Free-standing LAGLIDADG endonucleases typically recognize 18-22 bp (178). The intein-associated LAGLIDADG endonuclease PI-SceI, however, involves other parts of the intein in DNA binding (63) to recognize a bipartite 31 bp long sequence (59).

HNH and HisCys box

The HNH and HisCys box nucleases have very similar folds (called the ββα-Me motif). Based on these structural similarities it has been suggested that they should be considered as a singular group (96). **HNH** is the most diverse group of nucleases and can be divided into eight or more subgroups (115) depending on sequence variations in the core HNH motif and additional features. HNH homing endonucleases are usually found in phage introns, e.g. I-TevIII is located in the *nrdB* intron of phage RB3 (152). Typical members recognize asymmetric sequences that are 24 bp or longer, e.g. I-HmuI (61, 103). They are composed of a nuclease module that is quite non-specific and several modules that confer binding specificity. Many of them nick only one strand of the DNA (61, 103, 178).

Introns containing **HisCys box** endonucleases are hosted by genes in nuclear ribosomal DNA loci from several species of protists (e.g. I-PpoI (129) and I-DirI (80)). HisCys box endonucleases have a series of histidine and cysteine residues stretched over approximately 100 amino acids (81). The structure of I-PpoI shows that eight of the conserved residues coordinate a structurally important zink ion while others form the metal-coordinating active site (56, 57). HisCys box HEs recognize symmetric DNA sequences that are up to 20 bp long, with higher specificity than the HNH nucleases (178).

GIY-YIG

GIY-YIG homing endonucleases, like HNH homing endonucleases, are often found in phage introns, like the “paradigm” GIY-YIG nuclease I-TevI in the *td* intron of phage T4 (36, 197), but also in introns of fungal mitochon-

dria, algal mitochondria and algal chloroplasts and as freestanding ORFs like the T4 Seg endonucleases (167). Also like HNH homing endonucleases, GIY-YIG homing endonucleases are composed of modules, some of which they have in common, suggesting genetic exchanges between these two types of enzymes (178). Recognition sequences of the typical GIY-YIG HEs I-TevI and I-TevII are 37 and 31 bp long, respectively (10, 25, 36, 108).

PD-(D/E)-XK

This group of homing endonucleases, found in introns of cyanobacterial tRNA genes, was first believed to be an entirely new type of nuclease; only a few other ORFs were shown to be similar (15, 139). Many group I introns in bacteriophage contain protein encoding ORFs, but this was the "first example of a chromosomally encoded group I intron endonuclease in bacteria" (18). The group is represented by I-SspI, shown to cleave intron-less alleles of its host gene, recognizing 20 bp of partial symmetry (18). It has been shown now that this enzyme is a tetramer that conforms to the PD-(D/E)-XK fold commonly found in Type II restriction endonucleases (203).

Restriction endonucleases

Restriction modification systems (R-M) are thought to be an important defense for bacteria against invading DNA such as phage infections. By using a restriction endonuclease (RE) that is active only on modified DNA bacteria can specifically cleave modified DNA coming from the outside. Alternatively, by modifying their own DNA through methylation, bacteria are able to harbour an RE that only breaks down unmethylated DNA from the outside. There are four types of restriction endonucleases: Type I, Type II, Type III and Type IV (153), characterized by their different molecular structures, sequence recognition, cleavage position and cofactor requirements.

Type I

Type I REs are part of a restriction-modification system, usually with the composition R_2M_2S (*r*estriction, *m*ethylation and *s*pecificity subunits, respectively), where M_2S can perform the modification, methylating preferably at hemimethylated recongnition sites, while R_2M_2S is required for the cleavage reaction at unmethylated recognition sites (128), as for example for the well known EcoKI R-M system (48, 181). Recognition sites are bipartite and asymmetric, typically with 3-4 bases and 4-5 bases separated by a 6-8 base unspecified linker region (21, 154). Cleavage takes place at variable positions far outside the recognition sites, this requires ATP, S-adenosylmethionine (AdoMet) and Mg^{2+} (128) and occurs after two bound REs have dimerized and translocated DNA to produce a contracted loop that stalls the translocation (54).

Type III

Type III R-M systems are composed of a recognition-modification (Mod) and a restriction (Res) subunit, both of which are required for cleavage, while the Mod subunit is sufficient for hemimethylation of the recognition site (49). The stoichiometry of the subunits is $(Res)_2(Mod)_2$ (49, 77). Double-stranded cleavage by Type III REs is ATP- and Mg^{2+}-dependent, and is stimulated by AdoMet (e.g. EcoP15I (49, 125). The newly characterized PstII however, can use also GTP and CTP (165). Recognition sites are 5-6 bp long and asymmetric (21, 154). Efficient cleavage requires two enzymes bound to one site each in an inverse orientation and can be separated by thousands of basepairs (38, 49, 165). Similar to Type I REs, cleavage of unmethylated DNA is triggered by the collision of two complexes (21, 38). Type III endonucleases however, cleave at a fixed position close (25-27 bp) to one copy of the non-palindromic recognition sequence, as do the EcoP1I and EcoP15I enzymes (77, 125).

Type IV

The Type IV group of REs includes enzymes that require modified (methylated, hydroxymethylated or glucosyl-hydroxymethylated) DNA for cleavage and show weak specificity (153). REs of this group do not have an accompanying modification enzyme (49). The best example is the well studied McrBC from *E. coli* that recognizes a purine followed by a methylated or hydroxymethylated cytosine, but also is affected by surrounding sequences (147, 148, 176). Two sites that can be separated by 40-3000 bp are needed for cleavage, which takes place preferably around 30 bp away from one of the sites but also at a distance of 10, 20, 40 and 50 bp (21, 142, 177, 182). The structure of McrBC is very complex, the stoichiometry of the active complex bound to DNA (in the presence of GTP) is still not known. Initially however, McrB forms heptameric rings and tetradecamers which are stabilized by McrC (136, 142). A shorter form of the B subunit, which is unable to bind to DNA, is also expressed from the same gene at a 1:1 ratio and sequesters the catalytic C subunit (46, 138, 143). As for Type I and Type III REs Mg^{2+} is required, DNA translocation is GTP-dependent and cleavage is induced upon stalling of a complex (137, 182). Other examples are McrA (4, 147, 148), Mrr (26, 65, 87, 147, 190) and GmrSD (7, 8, 151), though these enzymes have not been as extensively characterized.

Type II

Type II REs is the most ubiquitous group of enzymes whith 3765 members listed in REBASE (155) as of October 29 2008 (http://rebase.neb.com). The majority of Type II restriction endonucleases contain the conserved PD-(D/E)-XK motif (144). A smaller fraction belong to the PLD, HNH or GIY-YIG families of endonucleases (e.g. BfiI, KpnI and Eco29kI, respectively)

(76, 162, 163). Unlike Type I, Type III and Type IV REs, most Type II REs function as freestanding enzymes that can cleave without being in complex with other subunits (144). They often bind as homodimers to their usually 4-8 bp long palindromic recognition sequences. Cleavage takes place on both strands within or close to this sequence and requires Mg^{2+}. REs with this typical behaviour are called **Type IIP** (*p*alindromic), typical examples are EcoRI (64) and EcoRV (164). Other subgroups of Type II REs are IIS, IIA, IIB, IIC, IIE, IIF, IIG, IIH, IIM and IIT (144, 153), but many REs can fit into more than one subgroup. Type **IIS** (*s*hifted) REs are enzymes that cleave outside their recognition sequence in at least one of the the two strands, although usually both strands are cleaved outside. FokI is the best studied Type IIS RE. It is a monomer in solution but is active as a dimer, interacting with two asymmetric recognition sites (16, 192) cleaving 9 and 13 bp away from one of the sites (179). Another TypeIIS RE that binds as a dimer to two DNA molecules is MboII (174), but there are also tetrameric Type IIS enzymes like BspMI (62).

Type **IIA** (*a*symmetric) REs recognize asymmetric sequences and cleave these either inside or outside the recognition sequence (144, 153). Those that cleave outside the recognition sequence thus also belong to the Type IIS group, like FokI (179). **Type IIB** REs are those that cleave on both sides of the recognition sequence, like BplI, which cleaves 13 and 8 bp (top and bottom strand respectively) before and 8 and 13 bp after the 11 bp long bipartite recognition sequence (199). **Type IIC** REs are enzymes that are fused to their modification domain, like BcgI, which is composed of two equal subunits containing the endonuclease/methyltransferase and one subunit containing the recognition domain (90, 91). This enzyme also cleaves on both sides of the recognition site (89), thus it also belongs to Type IIB REs. **Type IIE** REs interact with two target sites, cleaving only one of them, while the other site functions as an allosteric effector as has been shown for the well known EcoRII (141, 149) and NaeI (74). **Type IIF** REs also interact with two recognition sites, but unlike Type IIE cleave at both sites. REs of this group are homotetrameric like NgoMIV, which was the first tetrameric RE to be crystallized together with its substrate bound to two sites simultaneously (43, 55, 124), and SgrAI which tetramerizes upon binding to the DNA (14, 41, 70).

Type IIG REs, like Type IIC, have the restriction domain fused to the modification domain, but are also affected by AdoMet either positively or negatively. For example, the Eco57I R-M system consists of one enzyme with cleaving and methylating properties and a separate enzyme with methylating properties; cleavage is stimulated by AdoMet (78). Recognition sequences of Type IIG REs can be either symmetric or asymmetric. **Type IIH** REs are similar to Type I regarding gene structure. One of the few characterized enzymes of this group is AhdI, which has an accompanying methyltransferase consisting of two M and two S subunits (112). BcgI is only func-

tional as an A_2B complex where A contains both the restriction and methylation functions, while B contains the motifs necessary for recognition (91). **Type IIM** REs recognize and cleave methylated DNA at a fixed site. The best known example is DpnI, which cleaves the sequence GA↓TC where the A is methylated (101, 145). **Type IIT** REs are composed of heterodimers, like BslI which is believed to have the composition $\alpha_2\beta_2$ (73) and Bpu10I (175). REs can also be **nicking enzymes**, either naturally, like Nt.BstNBI (121) which nicks the top strand of the sequence $GAGTCN_4$↓ or due to mutations of e.g. one of the heterodimeric subunits of Type IIT REs like BsaI, BsmBI and BsmAI (204).

GIY-YIG endonucleases

Nucleases that share the GIY-YIG motif can be homing endonucleases, restriction endonucleases, excision repair nucleases, Penelope elements, MutS-like and Slx1 (50). The recognition of all the conserved motifs of GIY-YIG endonucleases is relatively recent, although Michel and Dujon noticed the sequence similarities between I-TevI and two other intron ORFs already in 1986 (116). In 1999 Kowalski *et al.* characterized the conserved sequence elements of the GIY-YIG module (Fig. 2) (94), which was later revised by Dunin-Horkawicz *et al.* in 2006 (50).

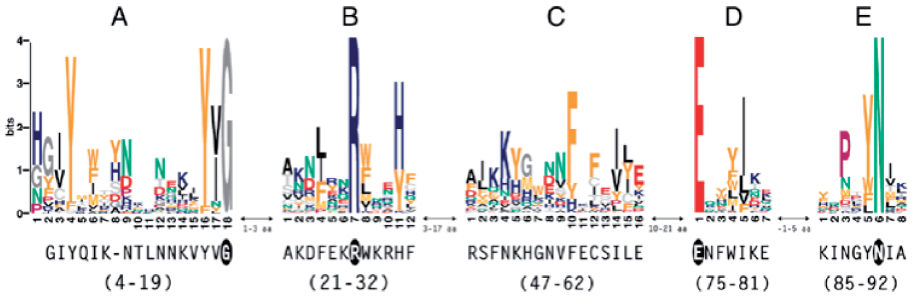

Figure 2. Sequence logo of conserved sequence motifs of the GIY-YIG module. The height of each amino acid is proportional to its conservation in that position. The sequence shown below is that of I-TevI. Reprinted from Kowalski *et al.* (94) with permission from the publisher.

Also in 1999, Aravind *et al.* recognized similarities between UvrC, intron-encoded endonucleases and several uncharacterized proteins and named the domain Uri (after *Uvr*C and *i*ntron-encoded endonucleases) (5). Hence the GIY-YIG domain is also called Uri. Dunin-Horkawicz *et al.* identified 765 database entries as GIY-YIG nucleases, from all domains of life, many of which have not been characterized (50). This analysis showed that the architecture of GIY-YIG nucleases is highly modular, the nuclease domain is usually attached to additional domains that confer DNA binding, protein

interactions, reverse transcription, additional catalysis and more. In 2002 the first structure of a GIY-YIG nuclease, I-TevI, was solved by Van Roey *et al.* (191). Also the structures of UvrC from both *Thermotoga maritima* and *Bacillus caldotenax* have been solved (188).

Examples of GIY-YIG endonucleases

The best characterized examples of GIY-YIG nucleases, except EndoII, are I-TevI and UvrC. The HE I-TevI is built up by two domains that are connected by a flexible linker which includes a Zn finger (42). The N-terminal domain harbours the catalytic GIY-YIG motif and the C-terminal domain confers most of the binding energy (24, 44, 94). I-TevI cleaves at a distance of 23 and 25 nucleotides (top and bottom strand respectively) from the recognition site (10). Due to the flexible linker, and the tolerance for sequence variations by the relatively unspecific catalytic domain, I-TevI can also cleave at a slightly different distance and can tolerate partial deletions of the flexible linker or the recognition sequence (24, 42). Nicking of the bottom strand is not Mg^{2+}-dependent, while nicking of the top strand is. The C-terminal domain of I-TevI is very similar to the C-terminal DNA binding domain of I-HmuI (178), emphasizing the modular architecture of GIY-YIG nucleases. The catalytic N-terminal domain of I-TevI has been extensively studied, showing that mutation of conserved residues of the GIY-YIG motif severely affects catalysis (94). The solved crystal structure showed that the conserved glutamic acid coordinates the Mg^{2+} ion that is essential for nicking of the top strand, and that also many of the other conserved residues are involved in forming the putative catalytic surface (191).

UvrC is an excision-repair nuclease which works in complex with UvrA and UvrB. The $UvrA_2UvrB_2$ complex is thought to scan the DNA until it finds a damage, UvrB binds to the damaged part while UvrA dissociates (111, 120, 196). Once UvrA has dissociated, UvrC is recruited to perform two incisions, first one incision 3 - 4 bp 3' of the damage, then another 7 bp 5' of the damage (105, 160, 195). After the incised oligonucleotide has been removed, DNA repair systems fill up and seal the gap (34, 75, 133).

Like I-TevI, UvrC is built up by more than one domain. The N-terminal half contains the GIY-YIG motif that catalyzes the 3' incision reaction and a UvrB-interacting region, while the C-terminal half confers DNA binding and performs the 5' incision (5, 84, 118, 119, 188). UvrC can also bind alone as a homotetramer to damaged and undamaged DNA with the same affinity, which is believed to be a way to concentrate the enzyme where it is needed *in vivo* (183). Also the structure of the GIY-YIG domain of UvrC showed that the conserved glutamic acid coordinates the essential Mg^{2+} ion, futhermore several other conserved amino acids were shown to be important for catalysis of the 3' incision (188).

In both I-TevI and UvrC the additional domains are necessary for stable binding to the recognition site, the N-terminal GIY-YIG nuclease domains are not sufficient for stable binding on their own (44, 188). However, overexpression of the cloned N-terminal domain of I-TevI appears to be toxic to *E. coli*, indicating that this domain binds well enough to cleave DNA *in vivo*.

The I-TevI isoschizomer I-BmoI is a two-domain HE very similar to I-TevI (53). Also for this enzyme only top strand nicking is Mg^{2+} dependent and the two incision reactions are sequential (35). It was suggested that the top and bottom strand incisions by these HEs could be catalysed by two different catalytic surfaces or through a significant metal-dependent reorganization of the catalytic surface.

Homology modelling of the single-domain GIY-YIG RE Eco29kI suggested that this enzyme has a fold similar to that of the N-terminal domains of UvrC and I-TevI (76). Mutational analysis confirmed the roles of various conserved amino acids in catalysis and two of them (H108 and N154) were also implicated in binding. This enzyme has been shown to be monomeric in solution (140) but its stoichiometry upon DNA binding has not been determined. The RE Cfr42I is an isoschizomer to Eco29kI, recognizing the symmetric sequence CCGC↓GG; the two enzymes also share 32% sequence identity (58). Cfr42I has been shown to be tetrameric in solution and has significantly higher catalytic activity when bound to two DNA substrates than when bound to only one (58). It can also function with a wide varitey of divalent metal ions.

Structure of GIY-YIG endonucleases

The structure of the catalytic GIY-YIG domain of I-TevI (191) revealed a novel α/β fold with a three-stranded anti-parallel β-sheet in the center flanked by three α-helices (Fig. 3). Amino acids implied in catalysis (94) form a shallow concave surface that would be able to accomodate the DNA substrate. The structures of the N-terminal GIY-YIG domains of UvrC from *Bacillus caldotenax* and *Thermotoga maritima* (Fig. 3) show very similar folds compared to I-TevI (188), though with some variation. The two UvrC nucleases e.g. have an additional structurally important α-helix (α1), which is absent in I-TevI. There are also differences between the UvrC nucleases, e.g. the helix α5 in UvrC from *Bacillus caldotenax* is absent in UvrC from *Thermotoga maritima* (and I-TevI) and is replaced by a loop region. However, highly conserved residues in the UvrC enzymes were found to be organized in a small patch and mutational analysis proved several of them to be important in catalysis. This corresponded well with the putative catalytic surface of I-TevI and strongly suggests that this is the catalytic site of all GIY-YIG enzymes.

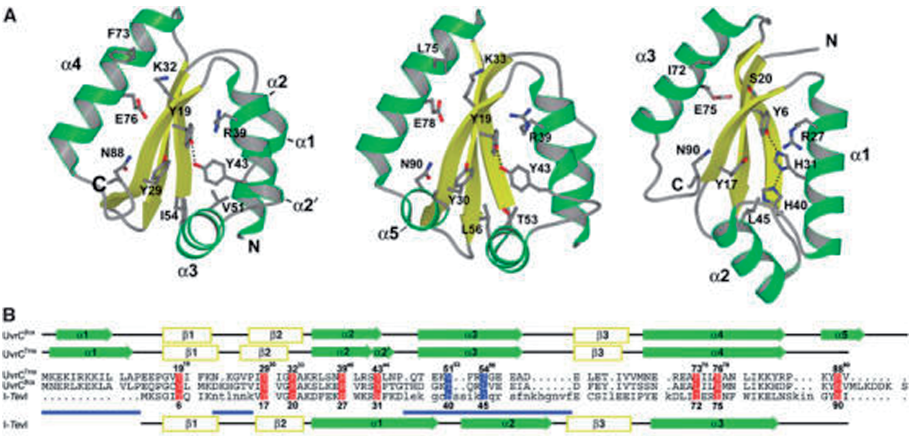

Figure 3. (A) Structures of the N-terminal catalytic GIY-YIG domains of UvrC from *Thermotoga maritima* (left), UvrC from *Bacillus caldotenax* (middle) and I-TevI (right). Dotted lines indicate hydrogen bonds. (B) Structure-based sequence alignment of the domains shown in A. Secondary structure elements are indicated above and below the sequence alignment corresponding to UvrC and I-TevI, respectively. The blue lines below the secondary structure elements indicate large regions of structural dissimilarity. Uppercase letters indicate residues that align structurally, while lowercase letters indicate residues that are not structurally aligned. Selected, structurally aligned residues are highlighted in red. Residues in a similar position, but not structurally aligned are highlighted in blue. Reprinted from Truglio *et al.* (188) with permission from the publisher.

GIY-YIG endonucleases encoded by bacteriophage T4

In addition to I-TevI and EndoII, T4 encodes several other GIY-YIG endonucleases. One is an intron-encoded homing endonuclease located in the self-splicing group I intron in *nrdD* (also called *sunY*) (9, 108). The SegA-G (*s*imilar to *e*ndonucleases of *g*roup I introns) enzymes also belong to the GIY-YIG group of nucleases. Sharma *et al*. found that the *segA* gene was homologous to I-TevI, and that the N-terminal 100 amino acids also were similar to those of four other uncharacterized proteins that were named SegB-E (167). SegA was shown to be Mg^{2+}-dependent and cut at preferred sites. It has not been proven to be a HE, but shows many similar properties. Since it is not positioned in an intron it was instead suggested that it could induce homing of the endonuclease-encoding gene (167). SegA is able to cleave cytosine-containing DNA as well as DNA containing glucosylated and hydroxymethylated cytosines with some sequence preference, and is stimulated by ATP (168). SegB has been shown to promote homing of its own coding sequence into related phage lacking the *segB* gene (22). SegC is also a site-specific endonuclease that is inherited by almost all phage in mixed infections with T4 and T2 and has been used for studying recombination events during T4 infections (169). Also SegE has been shown to pro-

mote transfer of its coding sequence to related phage lacking this sequence, and unpublished data report that SegD can initiate similar events (82). In mixed phage infections, the SegF and SegG nucleases promote the inheritance of their coding sequences and nearby regions to phage lacking the *seg* genes (11, 107). SegF shows sequence similarities in its C-terminal part to the HNH nuclease MobD. Another Seg nuclease, SegH, was found in phage T6, RB3 och LZ2, but not in T4 (161). In the genomes of five other T4-like phages only one similar gene (most similar to *segD*) was identified, hence GIY-YIG HE genes seem to be more abundant in T4 compared to related phage (130). It has been suggested that T4 perhaps is less sensitive to the action of these nucleases due to efficient DNA repair and therefore can harbour them in its genome (11). In addition, the concerted action of all T4 HEs might fragment the genomes of other phage in mixed infections, giving T4 the advantage.

EndoII

Endonuclease II, encoded by the gene *denA* (*D*NA *en*donuclease; (67)), is a GIY-YIG nuclease but does not fit strictly into any of the functional groups of nucleases. It behaves like a restriction endonuclease in that it cuts DNA foreign to itself. The "self DNA", however, is not protected by an accompanying methylase as is common for REs, but by hydroxymethylated cytosines that are incorporated into the T4 genome and subsequently glucosylated. The recognition sequence, however, is more like that of a homing endonuclease - ambiguous, long and asymmetric.

Microarray analysis of RNA transcripts during T4 infection has shown that *denA* is transcribed as immediate early (0-3 minutes) after infection (110). There is no obvious early promoter in the region before *denA* but it was suggested that there is an early promoter near the beginning of *nrdB* (110). Fragmentation of cytosine-containing phage DNA during infection with a T4 strain defective in hydroxymethylation and the 46/47 exonuclease was first studied by Kutter and Wiberg (100). Phage DNA thus was not protected from endonucleolytic cleavage by phage-encoded enzymes and cleavage products were not further degraded by exonucleases. A difference in fragmentation between phage DNA and host DNA was noted. Host DNA was cleaved to approximately 1.5 kDa fragments, while cytosine-containing phage DNA was cleaved to approximately 15 kDa fragments. It was suggested that this was dependent on a sequence-specific endonuclease with a preference for GC-rich sequences. EndoII was first named and purified from T4 infected *E. coli* and characterized *in vitro* in 1969, though the enzyme preparation was quite crude and most likely contaminated by EndoIV and possibly other enzymes (157). None-the-less it was determined to be a Mg^{2+}-dependent single-strand nicking enzyme, producing 3'-hydroxyl and 5'-phosphate termini of the DNA.

The first *in vivo* demonstration of discrete DNA fragments as the result of "restriction" was performed with EndoII (33). It was shown that the resulting fragments corresponded to genetically distinct regions of T4 DNA. A mutation located in *denA*, the gene encoding EndoII, was found to inhibit this degradation, showing that EndoII was required for this restriction. It was also the first actual evidence that EndoII was sequence specific.

Conclusive evidence that EndoII is the nuclease primarily responsible for the degradation of cytosine-containing DNA during T4 infection was presented by Carlson and Øvervatn (32). Degradation by EndoII was shown to be inhibited by as little as 3-9% hydroxymethylated cytosines in the DNA, corresponding to one hydroxymethylated cytosine every 30-95 bp. Studying the cleavage of plasmid pBR322 *in vivo*, the ambiguous consensus sequence 5'-CGRC**CG**CN↑TTGSYNGC-3' recognized by EndoII was determined (30). The most conserved parts of the sequence were in positions 4-7 (conserved sequence left, CSL) and positions 9-11 (conserved sequence right, CSR), though only the CG dinucleotide in positions 5 and 6 was completely conserved *in vivo*. Clevage of the bottom strand, i.e. the strand not shown, takes place between position 8 and 9, while cleavage of the top strand varies, generating blunt ends or 1-2-base 5' overhangs; a small number of single-stranded nicks was also observed. While showing that the 16 bp sequence was enough for cleavage, it was also found that the efficiency of *in vivo* cleavage was negatively influenced by the presense of additional preferred cleavage sites nearby (29). This context effect ranged from 800 to 1500 basepairs and it was suggested that EndoII engages more DNA than just the cleavage site. The context within the consensus sequence also influenced cleavage efficiency, supporting the idea that a structural motif was required for efficient cleavage. *In vitro* experiments, where EndoII was expressed through coupled *in vitro* transcription-translation from cloned *denA*, showed that EndoII recognized essentially the same sites *in vitro* as *in vivo* (28). Additional sites were also recognized and nicking was more frequent than double-strand cleavage. The consensus sequence as determined from *in vitro* experiments differed from that *in vivo* in that the CSR was no longer required.

Present investigation

Aim of the present study

The aim of the present study was to obtain a detailed understanding of how EndoII interacts with its DNA substrate. The main focus was on the following questions: why does cleavage by EndoII *in vivo* differ from that *in vitro* (paper I); what parts of the DNA substrate are recognized by EndoII and what amino acids are important for interactions with these (paper II); how does EndoII bind to its DNA substrate (paper III); what are the positions of the important amino acids in the three-dimensional structure of EndoII (paper V); is there an additional biological role for EndoII, besides host DNA degradation (paper IV). The results of these studies are presented below.

Structure of EndoII

The only available crystal structures of GIY-YIG nucleases are those of I-TevI and UvrC (from *Thermotoga maritima* and from *Fusobacterium nucleatum*) (188, 191). The crystal structure of EndoII has the same central fold as these, a three-stranded antiparallel β-sheet surrounded by helices, though with some variations (paper V). The structure revealed a tetrameric organization with an unusual means of dimerization and tetramerization. Half-moon shaped dimers are formed mainly by hydrophobic interactions between protruding β-sheet domains formed by β5 and β6 (Figs. 1 and 3 in paper V); there are no similar β-sheets present in the structures of the other GIY-YIG nucleases. The same domains also form the tetramerization surface, where two primary dimers bind back to back, resulting in a striking cross-like structure with two catalytic surfaces to each side of the cross. Each monomer contains all the amino acids to build up one catalytic surface. Gel filtration of EndoII mutants together with a 30 basepair DNA substrate showed that EndoII binds as a tetramer (paper III), hence the structure presented in paper V most likely is the biologically relevant structure.

In silico docking of a 16 basepair oligoduplex (from a favoured nick site) suggested that a loop containing a NUMOD3 DNA binding motif (170) could dock in the major groove of the DNA. This would position the scissile bond of the recognition sequence at the catalytic surface of the same monomer. For the other DNA strand to be able to contact the other catalytic surface of the primary dimer, severe distortion of the DNA would be necessary

since the second catalytic surface is placed relatively far away from the first. This might explain why single-stranded nicks is the most common outcome of EndoII activity. It is also possible that the C-terminal lysines are involved in DNA contacts to the right part of the recognition sequence.

EndoII binds as a tetramer to two DNA molecules

To elucidate the composition of enzyme-substrate complexes a combination of gel filtration and gel shift assays with enzymes and substrates of different lengths was performed (paper III). This allowed the identification of the number of DNA molecules and enzyme molecules in the different complexes formed. The complex formed first, at low enzyme concentrations, by all mutant enzymes except EndoII E118A (see below), was shown to consist of four EndoII molecules and two DNA molecules. The gel shift experiments also showed that a 30 bp substrate was not as efficiently bound as a 44 bp substrate by most mutants (E118A excepted also in this case), though they were nicked at the same positions. Thus, even though a 16 bp sequence is enough to direct nicking to a certain position, such a short substrate by itself is not enough for stable binding.

It is quite common that restriction enzymes form homodimers or tetramers to be able to execute double-stranded cleavage on a symmetric target site. For example the GIY-YIG RE Cfr42I has been shown to be a tetramer in solution and bind to two DNA molecules; binding of the second DNA molecule significantly increases cleavage activity (58). A tetrameric complex was quite unexpected for EndoII, however, since EndoII recognizes a long and asymmetric DNA sequence (paper II, (28-30)) with relatively low sequence specificity. Such sequence recognition is more similar to that of the double-strand-cutting homing endonucleases (I-TevI and I-BmoI) of the GIY-YIG family (25, 53) that bind as monomers.

Mutational analysis of EndoII

Other well characterized GIY-YIG endonucleases have additional domains that are responsible for the specific binding to their DNA substrates. The catalytic GIY-YIG domains of I-TevI, I-BmoI and UvrC can not bind strongly to DNA when separated from their DNA binding domain and mutants of these therefore can not be examined for altered binding properties (35, 44, 188). EndoII thus presents a unique opportunity to examine the roles of specific residues in the GIY-YIG module for binding and substrate specificity.

To this end conserved GIY-YIG residues in EndoII (G49, R57, E118 and N130) were selected for mutational analysis based on sequence similarities and available data on catalytic mutants of I-TevI and UvrC. Other suitable candidates for analysis were identified by aligning available sequences of 13

additional *denA* genes from T4-like phage with the T4 *denA* gene. Regions of homology not present in other GIY-YIG nucleases were identified in the middle region (MR) and the N-terminal region (NTR) of EndoII and selected for mutagenesis. The MR overlaps with the NUMOD3 DNA binding motif (170), which suggested that this region might be important for binding by EndoII. *denA* mutations found in T4 phage *in vivo* were also included in the analysis. All mutants were examined with regard to catalytic activity, binding efficiency and sequence recognition (paper II).

Conserved residues of the GIY-YIG motif

Mutation of completely conserved GIY-YIG residues G49, R57, E118 and N130 to alanine resulted in dramatically reduced catalytic activity (paper II, table 2). This is consistent with what has been found also for the paradigm GIY-YIG endonucleases I-TevI and UvrC.

E118

Mutation of E118 to alanine in EndoII rendered the enzyme inactive, although its binding affinity was among the highest for all mutants (paper II). In regard to binding however, the E118A mutant differed from all other mutants in that it formed complexes as a monomer with a single DNA molecule and a dimer with a single DNA molecule, as well as a tetramer with two DNA molecules as seen for all other mutants (paper III). The reason for this did not seem to lie in its multimerization capacities, since gel filtration in the absence of DNA clearly showed that the E118A mutant appeared as both a dimer and tetramer in solution, in the same proportions as for other mutants. Hence the answer most likely lies in a unique capacity to bind stably to the substrate as a monomer and dimer. Possibly, the replacement of the negative charge of the glutamic acid reduces the repulsion of the DNA phosphate backbone, alternatively, removal of the glutamic acid could result in reduction of some steric hindrance which would facilitate binding, or a combination of both.

E118 is positioned on the catalytic surface and the corresponding residues in I-TevI and UvrC have been shown to be Mg^{2+}-binding (188, 191). However, since the structure solved for EndoII is that of the E118A mutant a Mg^{2+}-coordinating function for this residue in EndoII can not be determined until a Mg^{2+}-binding mutant has been crystallized. On the other hand, the absence of activity for this mutant and the position of the substituting alanine on the active surface, overlapping that of I-TevI and UvrC, strongly indicate that E118 is the Mg^{2+}-binding residue also in EndoII.

R57

The R57A mutant was severely compromised in catalytic activity and in binding, though nicking was relatively more affected than binding compared

to other mutants (paper II). *In situ* activity assays also showed that a DNA substrate, which with all certainty was bound by EndoII R57A, was not efficiently nicked. Calculating the information content for the different positions of the sequences that R57A recognized (i.e. how conserved a certain base is in a certain position of the recognition site), showed a recognition pattern that differed from that of the wildtype enzyme and most other mutants (paper II). Recognition at the distal parts of the recognition sequence was lower, while a higher sequence conservation was observed close to the scissile bond with a novel preference for a G just to the left of this (Fig. 6 in paper II).

We suggested that for this mutant to be able to nick there is a novel requirement for intrinsic deformation of the DNA, implying that in the absence of the arginine the ability of the enzyme to induce DNA distortion is reduced. Thus R57 would be important for deformation of the substrate. The need for DNA distortion is supported by the structure of EndoII (paper V).

In the structure, R57 is positioned on the catalytic surface. It is possible that this residue has the same role in catalysis as that suggested for the corresponding residues in I-TevI and UvrC, stabilizing a pentacovalent reaction intermediate or stabilizing the negative charge of the leaving 5'-phosphate (188, 191); it might also be important for positioning the substrate correctly for nucleophilic attack.

G49

As for the R57A mutant the G49A mutant was severely compromised in nicking and binding activities and more affected in nicking than binding compared to other mutants (paper II). It also showed a somewhat similar variation in sequence recognition, suggesting that also this mutation introduces a need for intrinsic deformation of the DNA. It is not likely that the glycine would be directly involved in DNA binding or distortion; the structure of I-TevI shows that the conserved glycine residue lies just behind the Mg^{2+} ion and that there is little space to accommodate a larger residue in that position (188). G49 in EndoII occupies the same position (paper V); therefore the effect of the G49A mutation most likely depends on an introduced steric hindrance in accomodating the Mg^{2+} ion when the larger alanine is inserted. Supporting this, the G49A mutant nicked more efficiently with Mn^{2+} than with Mg^{2+}, suggesting that a smaller ion fits more easily when the relatively small glycine has been replaced by the bulkier alanine (paper II).

N130 and P127

The N130A mutant and the *in vivo* isolated P127L mutant were both severly affected in their catalytic abilities; the P127L mutant was also negatively affected in binding. Both mutants exhibited similar substrate recognition as the wildtype enzyme, but deviated from this and the other mutants in being relatively more active with Ca^{2+}, suggesting an increased space where the metal ion is bound. N130 is positioned in the catalytic surface very close to

E118, and P127 is very likely responsible for the fold in this region, affecting the position of N130. A proline corresponding to P127 is present also in UvrC and matches the structure of EndoII; it is not present in I-TevI however. The conserved asparagine in I-TevI has been proposed to have a structural role in the catalytic surface (191); for UvrC the asparagine has been suggested to position the catalytic domain of the enzyme correctly in relation to the other domains (188). From the structure of EndoII it seems likely that N130 has a structural function in stabilizing the C-terminus of EndoII, a small difference in the position of the C-terminus might block the active site from DNA binding.

Conserved residues in the NTR and MR

For MR and NTR mutants the binding abilities seemed to be more disturbed in relation to the catalytic abilities compared to what was seen for the other mutants (paper II, table 2). Both the MR and the NTR were thus suggested to contribute mainly to binding affinity (paper II), supported by the sequence similarity of parts of the MR to the NUMOD3 DNA binding motif (170). Mutants in the MR all behaved in a similar way with the exception of the K76A and L84P mutants which were as severely affected in catalysis as the GIY-YIG mutants.

K76

K76, which is part of the MR, is positioned in the catalytic surface. This residue is largely buried and hence not likely to take part in DNA binding; this is also supported by the relatively good binding ability of the K76A mutant. However, K76 most likely is involved in catalysis as reflected by the low catalytic activity of the K76A mutant enzyme. No corresponding residue is conserved in I-TevI or UvrC and there is no immediately corresponding residue occupying the same position in the otherwise so similar UvrC structure, suggesting that this is a unique catalytic feature of EndoII.

L84

L84P alone did not bind at all to the DNA substrate, hence it was also severely affected in catalysis (paper II). The total information content of the recognition sequence was also lower than that for any other mutant, reflecting a much lower precision in sequence recognition by the L84P mutant. L84P is positioned in the C-terminus of helix α3, which lies just before the NUMOD3 element. The severe defects in substrate binding by the L84P mutant is possibly caused by a disruption of the α3 helix, that is transmitted to the NUMOD3 motif.

Metal dependency of EndoII

Mg^{2+} is required for all EndoII cuts (paper I). This is in contrast to the GIY-YIG homing endonuclease I-TevI which nicks the two strands almost simultaneously, though only nicking of the second strand is Mg^{2+}-dependent, and also dependent on the first incision (126), suggesting different catalytic mechanisms for EndoII and I-TevI.

Replacing Mg^{2+} with Mn^{2+}, Ni^{2+} or Ca^{2+} in EndoII reduces its catalytic activity; the only exception is the G49A mutant that shows increased activity with Mn^{2+}. Galburt *et. al* proposed that a large effect of the metal species on catalysis may suggest that the metal ion is involved in positioning and activating a nucleophile for catalytic attack, rather than stabilizing the charge of the phosphoanion transition state (56). It is possible that this is the case for EndoII.

DNA repair masks nicking by EndoII *in vivo*

It was proposed that the differences between EndoII cleavage *in vitro* and *in vivo* are due to repair systems present in the *in vivo* situation (28). Therefore *in vitro* activity was assayed in the presence and absence of T4 DNA ligase to elucidate if DNA repair could be the reason for the different results (paper I). Cleavage of sites previously found to be favoured *in vitro* was significantly reduced when T4 DNA ligase was added. A site that was favoured *in vivo* however, was less affected by the addition of ligase. In order to test the kinetics with which the different sites were nicked or cleaved, upper and lower strands at three different *in vitro* favoured sites were assayed for nicking as well as for double-stranded cleavage by EndoII. Nicks appeared earlier and more abundantly than double-stranded cleavage, being at the very least three-fold higher than cleavage after five minutes. Even after twenty minutes cleavage never represented more than half of the least abundant nicks.

We concluded that double-stranded cleavage is the result of two separate nicking events, and that double-stranded cleavage *in vivo* will result only if these two events are close in time. Single-stranded nicks are thus masked *in vivo* by the activity of repair enzymes present during T4 infection of *E. coli*.

In vivo context effect

There are several examples of tetrameric enzymes binding to two DNA substrate molecules that loop the DNA between the two binding sites. One-dimensional diffusion along the DNA to find the second recognition site would explain the low amount of hydroxymethylation needed to inhibit EndoII activity on T4 phage DNA (one hydroxymethylated cytosine every 30-95 bp is enough to prevent cleavage *in vivo*) (32). Looping of the substrate bound to the two sites of the EndoII tetramer might explain the context ef-

fects seen *in vivo* (29), where cleavage of one site is reduced if a more preferred cleavage site is present within approximately 1000 bp to each side. If only one of the two sites bound by the EndoII tetramer is nicked or cleaved, this would explain the reduction of cleavage at the other, less preferred site. The absense of long range context effects *in vitro* can be explained by the fact that double-stranded cleavage *in vitro* often is the result of two separate nicking events; this is quite rare *in vivo* where both strands must be nicked almost simultaneously to persist.

EndoII can induce recombination events

As previously implicated (27) it seemed reasonable that EndoII could also have a second role during phage infection. This was based on the fact that EndoII *in vivo* produces fairly long (100-1000 bp) DNA fragments. This could be an incitement for the recombination of these gene sized fragments into the phage genome, thereby providing the phage with an opportunity to gain new functions.

To test this hypothesis isogenic recombinant T2 phage carrying the L84P mutation of T4 EndoII, or wildtype T4 EndoII, were grown in *E. coli* containing a plasmid with the T6 distal tail fiber gene *38*. Gene *38* is the gene conferring host range specificity since the gene product binds to receptors on the bacterial cell wall. Progeny phages were subsequently tested on T2-resistant hosts for their ability to grow on these. The frequency of recombinant host-range variants, i.e. phage progeny able to infect a T2-resistant host (readily infected by T6), was significantly higher for phage encoding functional EndoII. Phages able to grow on all resistant strains resulted from recombination at a few GC-rich positions in gene *38*, at locations with short sequence homology between the T2 and T6 genes (paper IV). These crossover regions regions were in the same multirecombination site that was described by Tétart *et al*. (184). It was also shown that EndoII cleaved the plasmid containing T6 gene *38* at and close to the sites of recombination *in vivo*; most likely nicking was more frequent than cleavage. EndoII thus increases the frequency of recombination events at sequences recognized by this enzyme and may be of evolutionary advantage to the phage.

Model for EndoII activity

Combining the results described above led to the following model for EndoII activity. Stable binding by EndoII is achieved when two substrate molecules are bound to the tetramer, one to each primary dimer. To find the second recognition site EndoII scans along the DNA by one-dimensional diffusion. Specific contacts to the DNA, close to the center of the consensus sequence, are made by the catalytic surface, while the distal parts of the DNA are contacted by the NUMOD3 loop. Possibly also the N-terminal lysines are in-

volved in DNA binding. In order for both NUMOD3 loops of a primary dimer to bind to the substrate, and for both catalytic surfaces to be able to make contact, the DNA must be considerably distorted, likely in part by residues of the catalytic surface.

The metal ion positions and activates a nucleophile for the catalytic attack and R57A stabilizes the pentacovalent reaction intermediate or, alternatively, stabilizes the 5' phosphate group. It is also possible that it positions the scissile bond correctly for nucleophilic attack. Due to the severe distortion required for double-stranded cleavage, the most common outcome is nicking of only one strand of the recognition site (mode I). However, if the DNA sequence is permissive (i.e. CSR is present) the second catalytic surface may also contact the DNA to effect double-stranded cleavage (mode II). *In vivo* single-stranded nicks are quickly repaired, so that mainly double-stranded cuts persist.

Concluding remarks

In conclusion, the results presented in this thesis have provided new insights into the function of GIY-YIG endonucleases in general and of EndoII in particular. It has been shown that catalytic residues of the conserved GIY-YIG motif are important for DNA recognition and binding in addition to catalysis. It also points to the diversity of GIY-YIG endonucleases in regard to biological role, cleavage specificity and structural organization outside the conserved GIY-YIG fold.

Future prospects

Further work on this enzyme that would be of interest is crystallization of EndoII in complex with DNA. This would give the first structure of a GIY-YIG endonuclease bound to its substrate. Comparison of co-crystal structures of for example the N130A, R57A, G49A and K76A mutants would directly show what amino acids interact with specific parts of the substrate and most likely give further clues to the reaction mechanism. The latter three mutants exhibit deviant sequence recognition while the N130A mutant is more wildtype-like in regard to recognition.

Since recognition by EndoII is so promiscuous it might also be possible to engineer a functional enzyme by fusing EndoII to a DNA binding domain, for example the DNA binding domain of I-TevI or I-BmoI. It would also be interesting to see if EndoII can bind to hydroxymethylated and glucosylated phage DNA as well as to host DNA, if so, the two types of DNA could be brought in close proximity in the same binding event to further promote recombination between the two.

Acknowledgements

Thank you,

Karin Carlson, my supervisor and friend, for supporting me during all these years, for inspiring discussions about science and family, our shared interest in Star Trek Voyager and a lot more ☺. I really appreciate that you have always taken the time to talk to me and answer my questions, even when you have been extremely busy teaching. I truly admire your efficiency and devotion to your work.

Evalena Andersson for teaching company in the beginning of time, scientific discussions with a lot of laughter, teaching me the ÄKTA system, for modeling of EndoII and for solving the structure! You're the best personal crystallographer I've ever had ☺.

Anna-Chey Nyström for introducing me to the labwork when I first came here and looong coffee-breaks ☺.

Patricia Lindgren you are the best of friends, always caring and ready to lend a helping hand, always honest and straight-forward which is something I wish more people could be. Thank you for kicking my butt once in a while ☺. **Oliver** for always calling me on my birthday and sending me that catalogue (you know which). **Vincent and Leandra** for all the nice birthday parties and playing in your garden, I hope you will love your new house and garden.
Fredrik Petterson, dear friend and my room-mate for a while, for great lab, children and party conversations. Soon it will be your turn - best of luck. **Janne Olsson**, my jujutsu instructor, dear friend and busbästis. You have a heart of gold (and a fist of steel). I hope you will always stay the same. **Helene Borg** for all the shared moments of family talks, coffee, cookies and candies and endless allegories - the number of crosses in the neuron network of your brain must be at least thousand-fold higher than in the average person. **Per Nilsson** for all the inspiring protein-related discussions, setting up the gel filtration SMART system, pleasant company, and for being a really nice guy. **Niklas Henriksson** also for protein-related discussions and nice company- you're next! After **Pontus Larsson** of course – good luck! **Shiying Wu**, I love your spirited way, and the amazing true stories from China

over sushi and beer. That smile of yours really lightens the mood on a gloomy day. **Maria Aveskog** for singing and smiling. **Siv Strömberg** for sequencing a long time ago. **Akiko** for always helping with computer stuff. **Ulrika Lustig** for your steady company through all these years. **Solveigh Ravnum** den "roliga" norskan, **Lotta Thuresson**, **Karin Hjort**- "the old gals". **Mathias Brännvall** (welcome back!), **Magnus Lindell** (good to have you here again), **Nisse Mikkelsen** - "the old guys". I really miss those good old times. **Fredrik Söderbom**, **Nina Grantcharova**, **Klas Flärdh**, **Mirjana McVanin**, **Ema Kikowska**. **Andrzej Poplawski** for salsa. **Klas Udekwu** for the company all those late nights in the lab. **Santanu Dasgupta** for volunteering to proofread my thesis and for nice traveling company on more than one occasion. **Bhupi Singh**, my latest room-mate and badminton coach. **Sonchita Bagchi** for that bright smile of yours. **Kristin Peisker** for nice badminton Sundays. **Cia Unosson, Aurelie Fender**, **Marie Wrande**, **Karin Troell -** the mommy-group and **Disa Hammarlöf** the soon-to-be-mommy. **Johan Reimegård, Erik Holmqvist**, **Tobias Norström**, **Nurul Islam** - the daddies. **Amanda Raine**, **Hava Lofton**, **Linda Marcusson** and **Mattias Andersson**, **Emma Ringqvist**, **Jon Jerlström Hultqvist**, **Henrik Tomenius**, **Jonas Lannergård**, **Lei Liu Conze**, **Nadja Heidrich**, **Maaike Sterk**, **Johan Ankarklev**. Good luck to you all! **Nora Ausmees** for shoe-shopping in Lund, the wonderful sushi dinner in Madison, the Art museum in Chicago and a lot more. **Diarmaid Hughes** for caring and always asking how things are going. All the other seniors, **Gerhart Wagner**, **Leif Kirsebom**, **Staffan Svärd**, **Anders Virtanen**, **Lars Hellman**, **Suparna Sanyal** and **Kurt Nordström**. Thank you **Eva Ferngren** for making a lot of the media and buffers I used, and for taking care of the coffee machine! Same to you **Ewa Gustafsson**. **Solan Rosén** for taking care of so many things. The administrators **Christina Friberger**, **Ingrid Söderström**, **Sigrid Pettersson**, **Anders Eriksson**, **Åsa Klefborg**, **Cristina Malm**, **Sofia Sundbaum** good luck with little Lisa. Thank you all for making ICM what it is, for nice company in the lunchroom, the lab and elsewhere!

To my wonderful family, I love you. **Mamma** and **Pappa** for always being there for me and encouraging me. My brothers **Rickard** for growing up with me and **Findus** for always fixing stuff for me (coffee, cars, dishwashers-you name it). Aunts, uncles and cousins and their children for all wonderful Christmases, summers in Finland etc, it is always nice to have you around. **Anita** and **Håkan** for being such nice in-laws. **Robert** and **Mia** and **Alva** for being you.

To the most important people in my life, **Fredrik**, **Miranda** and **Mirabel**, my beloved darlings forever.

Swedish summary

Endonukleas II - ett GIY-YIG enzym från bakteriofag T4

I det här arbetet har jag undersökt ett DNA-nedbrytande protein som tillverkas av bakteriofagen (fagen) T4. T4 infekterar *Escherichia coli* och andra enterobakterier genom att binda till deras cellvägg och föra in sitt eget DNA in i bakterien. Värdbakteriens proteiner kommer då att användas för att börja tillverka fagproteiner från fagens DNA. Efter 15-20 minuter vid 37˚C har alla komponenter som behövs för att tillverka ca 200 nya fagpartiklar tillverkats inuti bakterien, samtidigt som bakterien bryts ned inifrån. Proteinet som jag undersökt, endonukleas II (EndoII), har en del i detta i och med att det initierar nedbrytningen av bakteriens DNA. Fagens DNA är däremot skyddat från denna nedbrytning genom att den använder sig av modifierat DNA som är hydroxymetylerat i cytosinbaserna. EndoII bildas kort efter det att fagen fört in sitt DNA in i värdbakterien och börjar klippa upp dess DNA i ca 1000 baspar stora bitar. Därefter börjar även andra fagenzymer bryta ned bakterie-DNAt. På detta sätt medverkar EndoII till att fagen kan använda bakteriens nedbrutna DNA till att bygga nytt fagDNA för att tillverka nya fagpartiklar. Det skyddar även fagen mot vidare motattack från bakterien - när inget DNA finns kvar kan bakterien inte tillverka nya enzymer som eventuellt skulle kunna slå ut fagens attack.

Det jag undersökt närmare är de exakta verkningsmekanismerna hos det här enzymet. Som en del av detta har följande frågor ställts: varför skiljer sig nedbrytningen av DNAt åt beroende på om den görs *in vivo* (i bakterien) eller *in vitro* (i provröret) (artikel I); vilka DNA-sekvenser känner EndoII igen och hur påverkar olika mutationer i enzymet denna igenkänning (artikel II); hur binder EndoII till DNAt (artikel III); hur ser EndoII ut på molekylär nivå (artikel V); kan EndoII ha en annan, hittills okänd biologisk roll förutom att bryta ned bakterieDNA (artikel IV).

I den första artikeln visar vi att EndoII faktiskt är ett enzym som oftast bara klipper den ena tråden av DNAt och att anledningen till att det ser ut att vara mestadels dubbeltrådsbrott *in vivo* beror på att bakteriens enzymer snabbt kan laga de flesta enkeltrådsbrotten. Genom att tillsätta DNA ligas till *in vitro*-reaktionen kan *in vivo*-situationen efterhärmas. Det är alltså bara de enkeltrådsbrott som ligger väldigt nära varandra och som klyvs nästan samtidigt som kan resultera i dubbeltrådig klyvning i bakterien, medan det *in*

vitro inte finns några enzymer som kan laga enkeltrådsbrotten om man inte själv tillsätter dem. Detta för oss ett litet steg närmare hur enzymet fungerar.

EndoII tillhör en grupp av enzymer som kallas GIY-YIG enzymer efter det konserverade motiv av aminosyror som återfinns i alla enzymer i denna grupp. I den andra artikeln har ett stort antal mutanter konstruerats där vissa av dessa aminosyror, som genom jämförelser med liknande enzymer förutspåtts vara viktiga för enzymets funktion, bytts ut mot den relativt harmlösa aminosyran alanin. Även andra aminosyror, som är lika bara mellan de EndoII-homologer som kodas av T4-lika fager, har bytts ut. Dessa mutanta enzymer har sedan uttryckts i *E. coli*, renats och analyserats för att se om någon av dem skiljer sig från vildtypen (d.v.s. original-enzymet). Tyvärr är det inte möjligt att rena vildtypsenzymet på samma sätt eftersom det är alltför giftigt för *E. coli* och hinner bryta ned allt DNA innan det uttryckts i tillräckligt stora mängder. Därför uttrycks vildtypsenzymet inte från DNA i *E. coli* utan i vetegroddsextrakt direkt från mRNA som producerats separat. De egenskaper som jämförts är hur bra de olika mutanta enzymerna binder till DNAt, hur bra de klyver och om de klyver på samma ställe i DNAt. De flesta av mutanterna var mer eller mindre negativt påverkade i sin aktivitet på DNA. De mutanter som var mest negativt påverkade gällande klyvningen var de som utgör den förmodade katalytiska ytan: glutaminsyran i position 118 (E118A), glycinen i position 49 (G49A) och argininen i position 57 (R57A). Förutom dessa var en mutant med lysinen i position 76 utbytt mot alanin (K76A) nästan inaktiv. I liknande enzymer (t.ex. I-TevI och UvrC) förmodas en glutaminsyra motsvarande EndoIIs E118 vara den aminosyra som koordinerar den essentiella magnesiumjonen. E118A-mutanten var helt inaktiv, vilket överensstämmer med data från dessa andra enzymer. Vi kunde dessutom visa att denna mutant fortfarande binder bra till DNAt, men med ett annorlunda bindningsmönster jämfört med alla andra mutanter. Mutanterna G49A, R57A och K76A uppvisade även en något annorlunda preferens för olika klyvningssäten. Den låga aktiviteten av K76A, och den tredimensionella strukturen av EndoII som visar att den utgör en del av den aktiva ytan, visar att denna lysin är viktig för katalysen. Rollen till G49 tros vara att ge plats åt den essentiella magnesiumjon som koordineras av E118. R57 tros stabilisera en reaktionsintermediär. Dessutom verkar det som att G49 och R57, som även är negativt påverkade i bindning, har en roll i att böja DNAt för att detta ska kunna klyvas effektivt. Den aminoterminala delen av enzymet och en mittregion som bara visar sekvenslikheter mellan EndoII-lika enzymer, förmodas framförallt bidra till bindningsstyrka till DNA.

I den tredje artikeln analyserades enzymets bindning till DNA mer i detalj. Alla mutanter visade sig binda som tetramerer till två stycken DNA-molekyler, vilket tyder på att även vildtypsenzymet binder på detta sätt. En av mutanterna, E118A, uppvisade dock ett annorlunda bindningsmönster. Detta visade sig bero på att denna mutant kan binda stabilt även som mono-

mer och dimer till en DNA molekyl, vilket ingen annan av mutanterna kan göra.

I den fjärde artikeln framkommer en hittills outforskad biologisk roll för EndoII. Förutom att enzymet bryter ned bakteriens DNA under infektion så bidrar nedbrytningen också till en ökad rekombination i faggenomet. Detta sker förmodligen på grund av de genstora fragment som EndoII producerar när den bryter ned bakterieDNAt. Det vi tittat på är frekvensen av uppkomsten av fager som visar en ny värdspecificitet genom rekombination med en främmande svansfibergen placerad i en plasmid i värdcellen. För en fag med funktionellt EndoII ökar denna frekvens upp till 10 gånger jämfört med en fag med inaktivt EndoII.

I artikel V presenteras den tredimensionella strukturen för EndoII-mutanten E118A. Strukturen visar en tetramer med en slående X-form, där två förmodade katalytiska ytor sitter på var sida om X-et vända mot varandra, men något förskjutna i sidled. Det är den hittills fjärde strukturen i världen som blivit löst för den här sortens enzym och den första för ett protein där den katalytiska domänen självständigt binder stabilt till DNA. Tack vare den kan vi betydligt lättare förstå hur enzymet fungerar och relatera resultatet av mutationsanalysen till var de muterade aminosyrorna sitter i enzymet.

Kombinationen av den tredimensionella strukturen och de data vi fått fram dels genom mutationsanalyserna, dels genom analysen av hur enzymet binder, och insikten att EndoII kan öka rekombinationsfrekvensen in vivo har nu hjälpt oss att få en riktigt bra bild av hur just det här enzymet fungerar. Det är också ett stort steg framåt för att förstå hur andra enzymer i GIY-YIG-familjen fungerar.

References

1. **Aagaard, C., M. J. Awayez, and R. A. Garrett.** 1997. Profile of the DNA recognition site of the archaeal homing endonuclease I-DmoI. Nucleic Acids Res **25:**1523-30.
2. **Abedon, S. T.** 1994. Lysis and the Interaction between Free Phages and Infected Cells, p. 397-405. *In* J. D. karam (ed.), Molecular Biology of Bacteriophage T4. ASM Press, Washington D.C.
3. **Allet, B., and J. D. Rochaix.** 1979. Structure analysis at the ends of the intervening DNA sequences in the chloroplast 23S ribosomal genes of C. reinhardii. Cell **18:**55-60.
4. **Anton, B. P., and E. A. Raleigh.** 2004. Transposon-mediated linker insertion scanning mutagenesis of the Escherichia coli McrA endonuclease. J Bacteriol **186:**5699-707.
5. **Aravind, L., D. R. Walker, and E. V. Koonin.** 1999. Conserved domains in DNA repair proteins and evolution of repair systems. Nucleic Acids Res **27:**1223-42.
6. **Babic, A. C., E. J. Little, V. M. Manohar, J. Bitinaite, and N. C. Horton.** 2008. DNA distortion and specificity in a sequence-specific endonuclease. J Mol Biol **383:**186-204.
7. **Bair, C. L., and L. W. Black.** 2007. A type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs. J Mol Biol **366:**768-78.
8. **Bair, C. L., D. Rifat, and L. W. Black.** 2007. Exclusion of glucosyl-hydroxymethylcytosine DNA containing bacteriophages is overcome by the injected protein inhibitor IPI*. J Mol Biol **366:**779-89.
9. **Bell-Pedersen, D., S. Quirk, J. Clyman, and M. Belfort.** 1990. Intron mobility in phage T4 is dependent upon a distinctive class of endonucleases and independent of DNA sequences encoding the intron core: mechanistic and evolutionary implications. Nucleic Acids Res **18:**3763-70.
10. **Bell-Pedersen, D., S. M. Quirk, M. Bryk, and M. Belfort.** 1991. I-TevI, the endonuclease encoded by the mobile td intron, recognizes binding and cleavage domains on its DNA target. Proc Natl Acad Sci U S A **88:**7719-23.
11. **Belle, A., M. Landthaler, and D. A. Shub.** 2002. Intronless homing: site-specific endonuclease SegF of bacteriophage T4 mediates localized marker exclusion analogous to homing endonucleases of group I introns. Genes Dev **16:**351-62.

12. **Berg, O. G., R. B. Winter, and P. H. von Hippel.** 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. Biochemistry **20:**6929-48.
13. **Berger, H., A. J. Warren, and K. E. Fry.** 1969. Variations in genetic recombination due to amber mutations in T4D bacteriophage. J Virol **3:**171-5.
14. **Bilcock, D. T., L. E. Daniels, A. J. Bath, and S. E. Halford.** 1999. Reactions of type II restriction endonucleases with 8-base pair recognition sites. J Biol Chem **274:**36379-86.
15. **Biniszkiewicz, D., E. Cesnaviciene, and D. A. Shub.** 1994. Self-splicing group I intron in cyanobacterial initiator methionine tRNA: evidence for lateral transfer of introns in bacteria. Embo J **13:**4629-35.
16. **Bitinaite, J., D. A. Wah, A. K. Aggarwal, and I. Schildkraut.** 1998. FokI dimerization is required for DNA cleavage. Proc Natl Acad Sci U S A **95:**10570-5.
17. **Bonocora, R. P., G. Caignan, C. Woodrell, M. H. Werner, and D. M. Hinton.** 2008. A basic/hydrophobic cleft of the T4 activator MotA interacts with the C-terminus of E.coli sigma70 to activate middle gene transcription. Mol Microbiol **69:**331-43.
18. **Bonocora, R. P., and D. A. Shub.** 2001. A novel group I intron-encoded endonuclease specific for the anticodon region of tRNA(fMet) genes. Mol Microbiol **39:**1299-306.
19. **Bouet, J. Y., N. J. Campo, H. M. Krisch, and J. M. Louarn.** 1996. The effects on Escherichia coli of expression of the cloned bacteriophage T4 nucleoid disruption (ndd) gene. Mol Microbiol **20:**519-28.
20. **Bouet, J. Y., H. M. Krisch, and J. M. Louarn.** 1998. Ndd, the bacteriophage T4 protein that disrupts the Escherichia coli nucleoid, has a DNA binding activity. J Bacteriol **180:**5227-30.
21. **Bourniquel, A. A., and T. A. Bickle.** 2002. Complex restriction enzymes: NTP-driven molecular motors. Biochimie **84:**1047-59.
22. **Brok-Volchanskaya, V. S., F. A. Kadyrov, D. E. Sivogrivov, P. M. Kolosov, A. S. Sokolov, M. G. Shlyapnikov, V. M. Kryukov, and I. E. Granovsky.** 2008. Phage T4 SegB protein is a homing endonuclease required for the preferred inheritance of T4 tRNA gene region occurring in co-infection with a related phage. Nucleic Acids Res **36:**2094-105.
23. **Broker, T. R.** 1973. An electron microscopic analysis of pathways for bacteriophage T4 DNA recombination. J Mol Biol **81:**1-16.
24. **Bryk, M., M. Belisle, J. E. Mueller, and M. Belfort.** 1995. Selection of a remote cleavage site by I-tevI, the td intron-encoded endonuclease. J Mol Biol **247:**197-210.
25. **Bryk, M., S. M. Quirk, J. E. Mueller, N. Loizos, C. Lawrence, and M. Belfort.** 1993. The td intron endonuclease I-TevI makes extensive sequence-tolerant contacts across the minor groove of its DNA target. Embo J **12:**4040-1.

26. **Bujnicki, J. M., and L. Rychlewski.** 2001. Identification of a PD-(D/E)XK-like domain with a novel configuration of the endonuclease active site in the methyl-directed restriction enzyme Mrr and its homologs. Gene **267:**183-91.
27. **Carlson, K., and L. D. Kosturko.** 1998. Endonuclease II of coliphage T4: a recombinase disguised as a restriction endonuclease? Mol Microbiol **27:**671-6.
28. **Carlson, K., L. D. Kosturko, and A. C. Nystrom.** 1999. Sequence-specific cleavage by bacteriophage T4 endonuclease II in vitro. Mol Microbiol **31:**1395-405.
29. **Carlson, K., L. D. Kosturko, and A. C. Nystrom.** 1996. Short-range and long-range context effects on coliphage T4 endonuclease II-dependent restriction. J Bacteriol **178:**6419-26.
30. **Carlson, K., M. Krabbe, A. C. Nystrom, and L. D. Kosturko.** 1993. DNA determinants of restriction. Bacteriophage T4 endonuclease II-dependent cleavage of plasmid DNA in vivo. J Biol Chem **268:**8908-18.
31. **Carlson, K., and E. S. Miller.** 1994. General Procedures, p. 427-437. *In* J. D. Karam (ed.), Molecular Biology of Bacteriophage T4. ASM Press, Washington D.C.
32. **Carlson, K., and A. Overvatn.** 1986. Bacteriophage T4 endonucleases II and IV, oppositely affected by dCMP hydroxymethylase activity, have different roles in the degradation and in the RNA polymerase-dependent replication of T4 cytosine-containing DNA. Genetics **114:**669-85.
33. **Carlson, K., and J. S. Wiberg.** 1983. In vivo cleavage of cytosine-containing bacteriophage T4 DNA to genetically distinct, discretely sized fragments. J Virol **48:**18-30.
34. **Caron, P. R., S. R. Kushner, and L. Grossman.** 1985. Involvement of helicase II (uvrD gene product) and DNA polymerase I in excision mediated by the uvrABC protein complex. Proc Natl Acad Sci U S A **82:**4925-9.
35. **Carter, J. M., N. C. Friedrich, B. Kleinstiver, and D. R. Edgell.** 2007. Strand-specific Contacts and Divalent Metal Ion Regulate Double-strand Break Formation by the GIY-YIG Homing Endonuclease I-BmoI. J Mol Biol.**374**: 306-321
36. **Chu, F. K., F. Maley, A. M. Wang, J. Pedersen-Lane, and G. Maley.** 1991. Purification and substrate specificity of a T4 phage intron-encoded endonuclease. Nucleic Acids Res **19:**6863-9.
37. **Cowan, J., B. d'Acci, B. Guttman, and E. Kutter.** 1994. Gel analysis of T4 prereplicative proteins, p. 520-527. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, K. Carlson, and E. S. Miller (ed.), Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington DC.
38. **Crampton, N., S. Roes, D. T. Dryden, D. N. Rao, J. M. Edwardson, and R. M. Henderson.** 2007. DNA looping and translocation

provide an optimal cleavage mechanism for the type III restriction enzymes. Embo J **26:**3815-25.
39. **Crawford, J. T., and E. B. Goldberg.** 1980. The function of tail fibers in triggering baseplate expansion of bacteriophage T4. J Mol Biol **139:**679-90.
40. **Crowther, R. A., E. V. Lenk, Y. Kikuchi, and J. King.** 1977. Molecular reorganization in the hexagon to star transition of the baseplate of bacteriophage T4. J Mol Biol **116:**489-523.
41. **Daniels, L. E., K. M. Wood, D. J. Scott, and S. E. Halford.** 2003. Subunit assembly for DNA cleavage by restriction endonuclease SgrAI. J Mol Biol **327:**579-91.
42. **Dean, A. B., M. J. Stanger, J. T. Dansereau, P. Van Roey, V. Derbyshire, and M. Belfort.** 2002. Zinc finger as distance determinant in the flexible linker of intron endonuclease I-TevI. Proc Natl Acad Sci U S A **99:**8554-61.
43. **Deibert, M., S. Grazulis, G. Sasnauskas, V. Siksnys, and R. Huber.** 2000. Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. Nat Struct Biol **7:**792-9.
44. **Derbyshire, V., J. C. Kowalski, J. T. Dansereau, C. R. Hauer, and M. Belfort.** 1997. Two-domain structure of the td intron-encoded endonuclease I-TevI correlates with the two-domain configuration of the homing site. J Mol Biol **265:**494-506.
45. **Dickerson, R. E.** 1998. DNA bending: the prevalence of kinkiness and the virtues of normality. Nucleic Acids Res **26:**1906-26.
46. **Dila, D., E. Sutherland, L. Moran, B. Slatko, and E. A. Raleigh.** 1990. Genetic and sequence organization of the mcrBC locus of Escherichia coli K-12. J Bacteriol **172:**4888-900.
47. **Drivdahl, R. H., and E. M. Kutter.** 1990. Inhibition of transcription of cytosine-containing DNA in vitro by the alc gene product of bacteriophage T4. J Bacteriol **172:**2716-27.
48. **Dryden, D. T., L. P. Cooper, P. H. Thorpe, and O. Byron.** 1997. The in vitro assembly of the EcoKI type I DNA restriction/modification enzyme and its in vivo implications. Biochemistry **36:**1065-76.
49. **Dryden, D. T., N. E. Murray, and D. N. Rao.** 2001. Nucleoside triphosphate-dependent restriction enzymes. Nucleic Acids Res **29:**3728-41.
50. **Dunin-Horkawicz, S., M. Feder, and J. M. Bujnicki.** 2006. Phylogenomic analysis of the GIY-YIG nuclease superfamily. BMC Genomics **7:**98.
51. **Durrenberger, F., and J. D. Rochaix.** 1993. Characterization of the cleavage site and the recognition sequence of the I-CreI DNA endonuclease encoded by the chloroplast ribosomal intron of Chlamydomonas reinhardtii. Mol Gen Genet **236:**409-14.
52. **Eddy, S. R., and L. Gold.** 1991. The phage T4 nrdB intron: a deletion mutant of a version found in the wild. Genes Dev **5:**1032-41.

53. **Edgell, D. R., and D. A. Shub.** 2001. Related homing endonucleases I-BmoI and I-TevI use different strategies to cleave homologous recognition sites. Proc Natl Acad Sci U S A **98:**7898-903.
54. **Ellis, D. J., D. T. Dryden, T. Berge, J. M. Edwardson, and R. M. Henderson.** 1999. Direct observation of DNA translocation and cleavage by the EcoKI endonuclease using atomic force microscopy. Nat Struct Biol **6:**15-7.
55. **Embleton, M. L., V. Siksnys, and S. E. Halford.** 2001. DNA cleavage reactions by type II restriction enzymes that require two copies of their recognition sites. J Mol Biol **311:**503-14.
56. **Galburt, E. A., M. S. Chadsey, M. S. Jurica, B. S. Chevalier, D. Erho, W. Tang, R. J. Monnat, Jr., and B. L. Stoddard.** 2000. Conformational changes and cleavage by the homing endonuclease I-PpoI: a critical role for a leucine residue in the active site. J Mol Biol **300:**877-87.
57. **Galburt, E. A., B. Chevalier, W. Tang, M. S. Jurica, K. E. Flick, R. J. Monnat, Jr., and B. L. Stoddard.** 1999. A novel endonuclease mechanism directly visualized for I-PpoI. Nat Struct Biol **6:**1096-9.
58. **Gasiunas, G., G. Sasnauskas, G. Tamulaitis, C. Urbanke, D. Razaniene, and V. Siksnys.** 2008. Tetrameric restriction enzymes: expansion to the GIY-YIG nuclease family. Nucleic Acids Res **36:**938-49.
59. **Gimble, F. S., and J. Wang.** 1996. Substrate recognition and induced DNA distortion by the PI-SceI endonuclease, an enzyme generated by protein splicing. J Mol Biol **263:**163-80.
60. **Goldfarb, A., and P. Palm.** 1981. Control of promoter utilization by bacteriophage T4-induced modification of RNA polymerase alpha subunit. Nucleic Acids Res **9:**4863-78.
61. **Goodrich-Blair, H., and D. A. Shub.** 1996. Beyond homing: competition between intron endonucleases confers a selective advantage on flanking genetic markers. Cell **84:**211-21.
62. **Gormley, N. A., A. L. Hillberg, and S. E. Halford.** 2002. The type IIs restriction endonuclease BspMI is a tetramer that acts concertedly at two copies of an asymmetric DNA sequence. J Biol Chem **277:**4034-41.
63. **He, Z., M. Crist, H. Yen, X. Duan, F. A. Quiocho, and F. S. Gimble.** 1998. Amino acid residues in both the protein splicing and endonuclease domains of the PI-SceI intein mediate DNA binding. J Biol Chem **273:**4607-15.
64. **Hedgpeth, J., H. M. Goodman, and H. W. Boyer.** 1972. DNA nucleotide sequence restricted by the RI endonuclease. Proc Natl Acad Sci U S A **69:**3448-52.
65. **Heitman, J., and P. Model.** 1987. Site-specific methylases induce the SOS DNA repair response in Escherichia coli. J Bacteriol **169:**3243-50.

66. **Heller, K. J.** 1992. Molecular interaction between bacteriophage and the gram-negative cell envelope. Arch Microbiol **158:**235-48.
67. **Hercules, K., J. L. Munro, S. Mendelsohn, and J. S. Wiberg.** 1971. Mutants in a nonessential gene of bacteriophage T4 which are defective in the degradation of Escherichia coli deoxyribonucleic acid. J Virol **7:**95-105.
68. **Herendeen, D. R., G. A. Kassavetis, J. Barry, B. M. Alberts, and E. P. Geiduschek.** 1989. Enhancement of bacteriophage T4 late transcription by components of the T4 DNA replication apparatus. Science **245:**952-8.
69. **Herendeen, D. R., K. P. Williams, G. A. Kassavetis, and E. P. Geiduschek.** 1990. An RNA polymerase-binding protein that is required for communication between an enhancer and a promoter. Science **248:**573-8.
70. **Hingorani-Varma, K., and J. Bitinaite.** 2003. Kinetic analysis of the coordinated interaction of SgrAI restriction endonuclease with different DNA targets. J Biol Chem **278:**40392-9.
71. **Hirano, N., H. Ohshima, and H. Takahashi.** 2006. Biochemical analysis of the substrate specificity and sequence preference of endonuclease IV from bacteriophage T4, a dC-specific endonuclease implicated in restriction of dC-substituted T4 DNA synthesis. Nucleic Acids Res **34:**4743-51.
72. **Hosoda, J., E. Mathews, and B. Jansen.** 1971. Role of genes 46 and 47 in bacteriophage T4 reproduction. I. In vivo deoxyribonucleic acid replication. J Virol **8:**372-87.
73. **Hsieh, P. C., J. P. Xiao, D. O'Loane, and S. Y. Xu.** 2000. Cloning, expression, and purification of a thermostable nonhomodimeric restriction enzyme, BslI. J Bacteriol **182:**949-55.
74. **Huai, Q., J. D. Colandene, M. D. Topal, and H. Ke.** 2001. Structure of NaeI-DNA complex reveals dual-mode DNA recognition and complete dimer rearrangement. Nat Struct Biol **8:**665-9.
75. **Husain, I., B. Van Houten, D. C. Thomas, M. Abdel-Monem, and A. Sancar.** 1985. Effect of DNA polymerase I and DNA helicase II on the turnover rate of UvrABC excision nuclease. Proc Natl Acad Sci U S A **82:**6774-8.
76. **Ibryashkina, E. M., M. V. Zakharova, V. B. Baskunov, E. S. Bogdanova, M. O. Nagornykh, M. M. Den'mukhamedov, B. S. Melnik, A. Kolinski, D. Gront, M. Feder, A. S. Solonin, and J. M. Bujnicki.** 2007. Type II restriction endonuclease R.Eco29kI is a member of the GIY-YIG nuclease superfamily. BMC Struct Biol **7:**48.
77. **Janscak, P., U. Sandmeier, M. D. Szczelkun, and T. A. Bickle.** 2001. Subunit assembly and mode of DNA cleavage of the type III restriction endonucleases EcoP1I and EcoP15I. J Mol Biol **306:**417-31.
78. **Janulaitis, A., M. Petrusyte, Z. Maneliene, S. Klimasauskas, and V. Butkus.** 1992. Purification and properties of the Eco57I restric-

tion endonuclease and methylase--prototypes of a new class (type IV). Nucleic Acids Res **20:**6043-9.
79. **Jayaram, B., and T. Jain.** 2004. The role of water in protein-DNA recognition. Annu Rev Biophys Biomol Struct **33:**343-61.
80. **Johansen, S., M. Elde, A. Vader, P. Haugen, K. Haugli, and F. Haugli.** 1997. In vivo mobility of a group I twintron in nuclear ribosomal DNA of the myxomycete Didymium iridis. Mol Microbiol **24:**737-45.
81. **Johansen, S., T. M. Embley, and N. P. Willassen.** 1993. A family of nuclear homing endonucleases. Nucleic Acids Res **21:**4405.
82. **Kadyrov, F. A., M. G. Shlyapnikov, and V. M. Kryukov.** 1997. A phage T4 site-specific endonuclease, SegE, is responsible for a non-reciprocal genetic exchange between T-even-related phages. FEBS Lett **415:**75-80.
83. **Kanamaru, S., P. G. Leiman, V. A. Kostyuchenko, P. R. Chipman, V. V. Mesyanzhinov, F. Arisaka, and M. G. Rossmann.** 2002. Structure of the cell-puncturing device of bacteriophage T4. Nature **415:**553-7.
84. **Karakas, E., J. J. Truglio, D. Croteau, B. Rhau, L. Wang, B. Van Houten, and C. Kisker.** 2007. Structure of the C-terminal half of UvrC reveals an RNase H endonuclease domain with an Argonaute-like catalytic triad. Embo J **26:**613-22.
85. **Kashlev, M., E. Nudler, A. Goldfarb, T. White, and E. Kutter.** 1993. Bacteriophage T4 Alc protein: a transcription termination factor sensing local modification of DNA. Cell **75:**147-54.
86. **Kassavetis, G. A., and E. P. Geiduschek.** 1984. Defining a bacteriophage T4 late promoter: bacteriophage T4 gene 55 protein suffices for directing late promoter recognition. Proc Natl Acad Sci U S A **81:**5101-5.
87. **Kelleher, J. E., and E. A. Raleigh.** 1991. A novel activity in Escherichia coli K-12 that directs restriction of DNA modified at CG dinucleotides. J Bacteriol **173:**5220-3.
88. **Koch, T., A. Raudonikiene, K. Wilkens, and W. Ruger.** 1995. Overexpression, purification, and characterization of the ADP-ribosyltransferase (gpAlt) of bacteriophage T4: ADP-ribosylation of E. coli RNA polymerase modulates T4 "early" transcription. Gene Expr **4:**253-64.
89. **Kong, H., R. D. Morgan, R. E. Maunus, and I. Schildkraut.** 1993. A unique restriction endonuclease, BcgI, from Bacillus coagulans. Nucleic Acids Res **21:**987-91.
90. **Kong, H., S. E. Roemer, P. A. Waite-Rees, J. S. Benner, G. G. Wilson, and D. O. Nwankwo.** 1994. Characterization of BcgI, a new kind of restriction-modification system. J Biol Chem **269:**683-90.
91. **Kong, H., and C. L. Smith.** 1998. Does BcgI, a unique restriction endonuclease, require two recognition sites for cleavage? Biol Chem **379:**605-9.

92. **Kornberg, A., S. B. Zimmerman, S. R. Kornberg, and J. Josse.** 1959. Enzymatic Synthesis Of Deoxyribonucleic Acid. Influence Of Bacteriophage T2 On The Synthetic Pathway In Host Cells. Proc Natl Acad Sci U S A **45:**772-85.
93. **Kornberg, S. R., S. B. Zimmerman, and A. Kornberg.** 1961. Glucosylation of deoxyribonucleic acid by enzymes from bacteriophage-infected Escherichia coli. J Biol Chem **236:**1487-93.
94. **Kowalski, J. C., M. Belfort, M. A. Stapleton, M. Holpert, J. T. Dansereau, S. Pietrokovski, S. M. Baxter, and V. Derbyshire.** 1999. Configuration of the catalytic GIY-YIG domain of intron endonuclease I-TevI: coincidence of computational and molecular findings. Nucleic Acids Res **27:**2115-25.
95. **Kreuzer, K. N., and S. W. Morrical.** 1994. Initiation of DNA replication, p. 28-42. *In* J. D. Karam (ed.), Molecular Biology of Bacteriophage T4. ASM Press, Washington D.C.
96. **Kuhlmann, U. C., G. R. Moore, R. James, C. Kleanthous, and A. M. Hemmings.** 1999. Structural parsimony in endonuclease active sites: should the number of homing endonuclease families be redefined? FEBS Lett **463:**1-2.
97. **Kutter, E., R. Drivdahl, and K. Rand.** 1984. Identification and characterization of the alc gene product of bacteriophage T4. Genetics **108:**291-304.
98. **Kutter, E., T. White, M. Kashlev, M. Uzan, J. McKinney, and B. Guttman.** 1994. Effects on host Genome Structure and Expression, p. 357-368. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, and K. Carlson (ed.), Molecular Biology of bacteriophage T4. American Society for Microbiology, Washington DC.
99. **Kutter, E. M., D. Bradley, R. Schenck, B. S. Guttman, and R. Laiken.** 1981. Bacteriophage T4 alc gene product: general inhibitor of transcription from cytosine-containing DNA. J Virol **40:**822-9.
100. **Kutter, E. M., and J. S. Wiberg.** 1968. Degradation of cytosin-containing bacterial and bacteriophage DNA after infection of Escherichia coli B with bacteriophage T4D wild type and with mutants defective in genes 46, 47 and 56. J Mol Biol **38:**395-411.
101. **Lacks, S., and B. Greenberg.** 1975. A deoxyribonuclease of Diplococcus pneumoniae specific for methylated DNA. J Biol Chem **250:**4060-66.
102. **Lambert, A. R., D. Sussman, B. Shen, R. Maunus, J. Nix, J. Samuelson, S. Y. Xu, and B. L. Stoddard.** 2008. Structures of the rare-cutting restriction endonuclease NotI reveal a unique metal binding fold involved in DNA binding. Structure **16:**558-69.
103. **Landthaler, M., B. W. Shen, B. L. Stoddard, and D. A. Shub.** 2006. I-BasI and I-HmuI: two phage intron-encoded endonucleases with homologous DNA recognition sequences but distinct DNA specificities. J Mol Biol **358:**1137-51.

104. **Lejeune, D., N. Delsaux, B. Charloteaux, A. Thomas, and R. Brasseur.** 2005. Protein-nucleic acid recognition: statistical analysis of atomic interactions and influence of DNA structure. Proteins **61:**258-71.
105. **Lin, J. J., and A. Sancar.** 1992. Active site of (A)BC excinuclease. I. Evidence for 5' incision by UvrC through a catalytic site involving Asp399, Asp438, Asp466, and His538 residues. J Biol Chem **267:**17688-92.
106. **Little, E. J., and N. C. Horton.** 2005. DNA-induced conformational changes in type II restriction endonucleases: the structure of unliganded HincII. J Mol Biol **351:**76-88.
107. **Liu, Q., A. Belle, D. A. Shub, M. Belfort, and D. R. Edgell.** 2003. SegG endonuclease promotes marker exclusion and mediates co-conversion from a distant cleavage site. J Mol Biol **334:**13-23.
108. **Loizos, N., G. H. Silva, and M. Belfort.** 1996. Intron-encoded endonuclease I-TevII binds across the minor groove and induces two distinct conformational changes in its DNA substrate. J Mol Biol **255:**412-24.
109. **Lu, M. J., and U. Henning.** 1992. Lysis protein T of bacteriophage T4. Mol Gen Genet **235:**253-8.
110. **Luke, K., A. Radek, X. Liu, J. Campbell, M. Uzan, R. Haselkorn, and Y. Kogan.** 2002. Microarray analysis of gene expression during bacteriophage T4 infection. Virology **299:**182-91.
111. **Malta, E., G. F. Moolenaar, and N. Goosen.** 2007. Dynamics of the UvrABC nucleotide excision repair proteins analyzed by fluorescence resonance energy transfer. Biochemistry **46:**9080-8.
112. **Marks, P., J. McGeehan, G. Wilson, N. Errington, and G. Kneale.** 2003. Purification and characterisation of a novel DNA methyltransferase, M.AhdI. Nucleic Acids Res **31:**2803-10.
113. **Marshall, P., and C. Lemieux.** 1991. Cleavage pattern of the homing endonuclease encoded by the fifth intron in the chloroplast large subunit rRNA-encoding gene of Chlamydomonas eugametos. Gene **104:**241-5.
114. **Marshall, P., and C. Lemieux.** 1992. The I-CeuI endonuclease recognizes a sequence of 19 base pairs and preferentially cleaves the coding strand of the Chlamydomonas moewusii chloroplast large subunit rRNA gene. Nucleic Acids Res **20:**6401-7.
115. **Mehta, P., K. Katta, and S. Krishnaswamy.** 2004. HNH family subclassification leads to identification of commonality in the His-Me endonuclease superfamily. Protein Sci **13:**295-300.
116. **Michel, F., and B. Dujon.** 1986. Genetic exchanges between bacteriophage T4 and filamentous fungi? Cell **46:**323.
117. **Miller, E. S., E. Kutter, G. Mosig, F. Arisaka, T. Kunisawa, and W. Ruger.** 2003. Bacteriophage T4 genome. Microbiol Mol Biol Rev **67:**86-156.
118. **Moolenaar, G. F., K. L. Franken, D. M. Dijkstra, J. E. Thomas-Oates, R. Visse, P. van de Putte, and N. Goosen.** 1995. The C-

terminal region of the UvrB protein of Escherichia coli contains an important determinant for UvrC binding to the preincision complex but not the catalytic site for 3'-incision. J Biol Chem **270:**30508-15.

119. **Moolenaar, G. F., K. L. Franken, P. van de Putte, and N. Goosen.** 1997. Function of the homologous regions of the Escherichia coli DNA excision repair proteins UvrB and UvrC in stabilization of the UvrBC-DNA complex and in 3'-incision. Mutat Res **385:**195-203.
120. **Moolenaar, G. F., M. Schut, and N. Goosen.** 2005. Binding of the UvrB dimer to non-damaged and damaged DNA: residues Y92 and Y93 influence the stability of both subunits. DNA Repair (Amst) **4:**699-713.
121. **Morgan, R. D., C. Calvet, M. Demeter, R. Agra, and H. Kong.** 2000. Characterization of the specific DNA nicking activity of restriction endonuclease N.BstNBI. Biol Chem **381:**1123-5.
122. **Morton, D., E. M. Kutter, and B. S. Guttman.** 1978. Synthesis of T4 DNA and bacteriophage in the absence of dCMP hydroxymethylase. J Virol **28:**262-9.
123. **Mosig, G., and D. H. Hall.** 1994. Gene Expression: A Paradigm of Integrated Circuits, p. 127-131. *In* J. D. Karam (ed.), Molecular Biology of Bacteriophage T4. ASM Press, Washington D.C.
124. **Mucke, M., D. H. Kruger, and M. Reuter.** 2003. Diversity of type II restriction endonucleases that require two DNA recognition sites. Nucleic Acids Res **31:**6079-84.
125. **Mucke, M., S. Reich, E. Moncke-Buchner, M. Reuter, and D. H. Kruger.** 2001. DNA cleavage by type III restriction-modification enzyme EcoP15I is independent of spacer distance between two head to head oriented recognition sites. J Mol Biol **312:**687-98.
126. **Mueller, J. E., D. Smith, M. Bryk, and M. Belfort.** 1995. Intron-encoded endonuclease I-TevI binds as a monomer to effect sequential cleavage via conformational changes in the td homing site. Embo J **14:**5724-35.
127. **Mukai, F., G. Streisinger, and B. Miller.** 1967. The mechanism of lysis in phage T4-infected cells. Virology **33:**398-404.
128. **Murray, N. E.** 2000. Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). Microbiol Mol Biol Rev **64:**412-34.
129. **Muscarella, D. E., E. L. Ellison, B. M. Ruoff, and V. M. Vogt.** 1990. Characterization of I-Ppo, an intron-encoded endonuclease that mediates homing of a group I intron in the ribosomal DNA of Physarum polycephalum. Mol Cell Biol **10:**3386-96.
130. **Nolan, J. M., V. Petrov, C. Bertrand, H. M. Krisch, and J. D. Karam.** 2006. Genetic diversity among five T4-like bacteriophages. Virol J **3:**30.
131. **Nossal, N. G.** 1994. The Bacteriophage T4 DNA repllication fork, p. 43-53. *In* J. D. Karam (ed.), Molecular Biology of Bacteriophage T4. ASM Press, Washington D.C.

132. **Ohshima, H., N. Hirano, and H. Takahashi.** 2007. A hexanucleotide sequence (dC1-dC6 tract) restricts the dC-specific cleavage of single-stranded DNA by endonuclease IV of bacteriophage T4. Nucleic Acids Res **35:**6681-9.
133. **Orren, D. K., C. P. Selby, J. E. Hearst, and A. Sancar.** 1992. Post-incision steps of nucleotide excision repair in Escherichia coli. Disassembly of the UvrBC-DNA complex by helicase II and DNA polymerase I. J Biol Chem **267:**780-8.
134. **Orsini, G., M. Ouhammouch, J. P. Le Caer, and E. N. Brody.** 1993. The asiA gene of bacteriophage T4 codes for the anti-sigma 70 protein. J Bacteriol **175:**85-93.
135. **Pande, S., A. Makela, S. L. Dove, B. E. Nickels, A. Hochschild, and D. M. Hinton.** 2002. The bacteriophage T4 transcription activator MotA interacts with the far-C-terminal region of the sigma70 subunit of Escherichia coli RNA polymerase. J Bacteriol **184:**3957-64.
136. **Panne, D., S. A. Muller, S. Wirtz, A. Engel, and T. A. Bickle.** 2001. The McrBC restriction endonuclease assembles into a ring structure in the presence of G nucleotides. Embo J **20:**3210-7.
137. **Panne, D., E. A. Raleigh, and T. A. Bickle.** 1999. The McrBC endonuclease translocates DNA in a reaction dependent on GTP hydrolysis. J Mol Biol **290:**49-60.
138. **Panne, D., E. A. Raleigh, and T. A. Bickle.** 1998. McrBs, a modulator peptide for McrBC activity. Embo J **17:**5477-83.
139. **Paquin, B., S. D. Kathe, S. A. Nierzwicki-Bauer, and D. A. Shub.** 1997. Origin and evolution of group I introns in cyanobacterial tRNA genes. J Bacteriol **179:**6798-806.
140. **Pertzev, A. V., A. N. Kravetz, S. G. Mayorov, M. V. Zakharova, and A. S. Solonin.** 1997. Isolation of a strain overproducing endonuclease Eco29kI: purification and characterization of the homogeneous enzyme. Biochemistry (Mosc) **62:**732-41.
141. **Petrauskene, O. V., O. V. Babkina, V. N. Tashlitsky, G. M. Kazankov, and E. S. Gromova.** 1998. EcoRII endonuclease has two identical DNA-binding sites and cleaves one of two co-ordinated recognition sites in one catalytic event. FEBS Lett **425:**29-34.
142. **Pieper, U., D. H. Groll, S. Wunsch, F. U. Gast, C. Speck, N. Mucke, and A. Pingoud.** 2002. The GTP-dependent restriction enzyme McrBC from Escherichia coli forms high-molecular mass complexes with DNA and produces a cleavage pattern with a characteristic 10-base pair repeat. Biochemistry **41:**5245-54.
143. **Pieper, U., and A. Pingoud.** 2002. A mutational analysis of the PD.D/EXK motif suggests that McrC harbors the catalytic center for DNA cleavage by the GTP-dependent restriction enzyme McrBC from Escherichia coli. Biochemistry **41:**5236-44.
144. **Pingoud, A., M. Fuxreiter, V. Pingoud, and W. Wende.** 2005. Type II restriction endonucleases: structure and mechanism. Cell Mol Life Sci **62:**685-707.

145. **Pingoud, A., and A. Jeltsch.** 2001. Structure and function of type II restriction endonucleases. Nucleic Acids Res **29:**3705-27.
146. **Pirhonen, M., P. Heino, I. Helander, P. Harju, and E. T. Palva.** 1988. Bacteriophage T4 resistant mutants of the plant pathogen Erwinia carotovora. Microb Pathog **4:**359-67.
147. **Raleigh, E. A., R. Trimarchi, and H. Revel.** 1989. Genetic and physical mapping of the mcrA (rglA) and mcrB (rglB) loci of Escherichia coli K-12. Genetics **122:**279-96.
148. **Raleigh, E. A., and G. Wilson.** 1986. Escherichia coli K-12 restricts DNA containing 5-methylcytosine. Proc Natl Acad Sci U S A **83:**9070-4.
149. **Reuter, M., D. Kupper, A. Meisel, C. Schroeder, and D. H. Kruger.** 1998. Cooperative binding properties of restriction endonuclease EcoRII with DNA recognition sites. J Biol Chem **273:**8294-300.
150. **Riede, I.** 1987. Receptor specificity of the short tail fibres (gp12) of T-even type Escherichia coli phages. Mol Gen Genet **206:**110-5.
151. **Rifat, D., N. T. Wright, K. M. Varney, D. J. Weber, and L. W. Black.** 2008. Restriction endonuclease inhibitor IPI* of bacteriophage T4: a novel structure for a dedicated target. J Mol Biol **375:**720-34.
152. **Robbins, J. B., M. Stapleton, M. J. Stanger, D. Smith, J. T. Dansereau, V. Derbyshire, and M. Belfort.** 2007. Homing endonuclease I-TevIII: dimerization as a means to a double-strand break. Nucleic Acids Res **35:**1589-600.
153. **Roberts, R. J., M. Belfort, T. Bestor, A. S. Bhagwat, T. A. Bickle, J. Bitinaite, R. M. Blumenthal, S. Degtyarev, D. T. Dryden, K. Dybvig, K. Firman, E. S. Gromova, R. I. Gumport, S. E. Halford, S. Hattman, J. Heitman, D. P. Hornby, A. Janulaitis, A. Jeltsch, J. Josephsen, A. Kiss, T. R. Klaenhammer, I. Kobayashi, H. Kong, D. H. Kruger, S. Lacks, M. G. Marinus, M. Miyahara, R. D. Morgan, N. E. Murray, V. Nagaraja, A. Piekarowicz, A. Pingoud, E. Raleigh, D. N. Rao, N. Reich, V. E. Repin, E. U. Selker, P. C. Shaw, D. C. Stein, B. L. Stoddard, W. Szybalski, T. A. Trautner, J. L. Van Etten, J. M. Vitor, G. G. Wilson, and S. Y. Xu.** 2003. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res **31:**1805-12.
154. **Roberts, R. J., and D. Macelis.** 1997. REBASE-restriction enzymes and methylases. Nucleic Acids Res **25:**248-62.
155. **Roberts, R. J., T. Vincze, J. Posfai, and D. Macelis.** 2007. REBASE--enzymes and genes for DNA restriction and modification. Nucleic Acids Res **35:**D269-70.
156. **Rossmann, M. G., V. V. Mesyanzhinov, F. Arisaka, and P. G. Leiman.** 2004. The bacteriophage T4 DNA injection machine. Curr Opin Struct Biol **14:**171-80.

157. **Sadowski, P. D., and J. Hurwitz.** 1969. Enzymatic breakage of deoxyribonucleic acid. I. Purification and properties of endonuclease II from T4 phage-infected Escherichia coli. J Biol Chem **244:**6182-91.
158. **Sadowski, P. D., and J. Hurwitz.** 1969. Enzymatic breakage of deoxyribonucleic acid. II. Purification and properties of endonuclease IV from T4 phage-infected Escherichia coli. J Biol Chem **244:**6192-8.
159. **Sadowski, P. D., H. R. Warner, K. Hercules, J. L. Munro, S. Mendelsohn, and J. S. Wiberg.** 1971. Mutants of bacteriophage T4 defective in the induction of T4 endonuclease II. J Biol Chem **246:**3431-3.
160. **Sancar, A., and W. D. Rupp.** 1983. A novel repair enzyme: UVRABC excision nuclease of Escherichia coli cuts a DNA strand on both sides of the damaged region. Cell **33:**249-60.
161. **Sandegren, L., D. Nord, and B. M. Sjoberg.** 2005. SegH and Hef: two novel homing endonucleases whose genes replace the mobC and mobE genes in several T4-related phages. Nucleic Acids Res **33:**6203-13.
162. **Sapranauskas, R., G. Sasnauskas, A. Lagunavicius, G. Vilkaitis, A. Lubys, and V. Siksnys.** 2000. Novel subtype of type IIs restriction enzymes. BfiI endonuclease exhibits similarities to the EDTA-resistant nuclease Nuc of Salmonella typhimurium. J Biol Chem **275:**30878-85.
163. **Saravanan, M., J. M. Bujnicki, I. A. Cymerman, D. N. Rao, and V. Nagaraja.** 2004. Type II restriction endonuclease R.KpnI is a member of the HNH nuclease superfamily. Nucleic Acids Res **32:**6129-35.
164. **Schildkraut, I., C. D. Banner, C. S. Rhodes, and S. Parekh.** 1984. The cleavage site for the restriction endonuclease EcoRV is 5'-GAT/ATC-3'. Gene **27:**327-9.
165. **Sears, A., L. J. Peakman, G. G. Wilson, and M. D. Szczelkun.** 2005. Characterization of the Type III restriction endonuclease PstII from Providencia stuartii. Nucleic Acids Res **33:**4775-87.
166. **Severinova, E., K. Severinov, D. Fenyo, M. Marr, E. N. Brody, J. W. Roberts, B. T. Chait, and S. A. Darst.** 1996. Domain organization of the Escherichia coli RNA polymerase sigma 70 subunit. J Mol Biol **263:**637-47.
167. **Sharma, M., R. L. Ellis, and D. M. Hinton.** 1992. Identification of a family of bacteriophage T4 genes encoding proteins similar to those present in group I introns of fungi and phage. Proc Natl Acad Sci U S A **89:**6658-62.
168. **Sharma, M., and D. M. Hinton.** 1994. Purification and characterization of the SegA protein of bacteriophage T4, an endonuclease related to proteins encoded by group I introns. J Bacteriol **176:**6439-48.

169. **Shcherbakov, V., I. Granovsky, L. Plugina, T. Shcherbakova, S. Sizova, K. Pyatkov, M. Shlyapnikov, and O. Shubina.** 2002. Focused genetic recombination of bacteriophage t4 initiated by double-strand breaks. Genetics **162:**543-56.
170. **Sitbon, E., and S. Pietrokovski.** 2003. New types of conserved sequence domains in DNA-binding regions of homing endonucleases. Trends Biochem Sci **28:**473-7.
171. **Snyder, L., L. Gold, and E. Kutter.** 1976. A gene of bacteriophage T4 whose product prevents true late transcription on cytosine-containing T4 DNA. Proc Natl Acad Sci U S A **73:**3098-102.
172. **Snyder, L., and L. Jorissen.** 1988. Escherichia coli mutations that prevent the action of the T4 unf/alc protein map in an RNA polymerase gene. Genetics **118:**173-80.
173. **Sommer, N., V. Salniene, E. Gineikiene, R. Nivinskas, and W. Ruger.** 2000. T4 early promoter strength probed in vivo with unribosylated and ADP-ribosylated Escherichia coli RNA polymerase: a mutation analysis. Microbiology **146 (Pt 10):**2643-53.
174. **Soundararajan, M., Z. Chang, R. D. Morgan, P. Heslop, and B. A. Connolly.** 2002. DNA binding and recognition by the IIs restriction endonuclease MboII. J Biol Chem **277:**887-95.
175. **Stankevicius, K., A. Lubys, A. Timinskas, D. Vaitkevicius, and A. Janulaitis.** 1998. Cloning and analysis of the four genes coding for Bpu10I restriction-modification enzymes. Nucleic Acids Res **26:**1084-91.
176. **Stewart, F. J., D. Panne, T. A. Bickle, and E. A. Raleigh.** 2000. Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. J Mol Biol **298:**611-22.
177. **Stewart, F. J., and E. A. Raleigh.** 1998. Dependence of McrBC cleavage on distance between recognition elements. Biol Chem **379:**611-6.
178. **Stoddard, B. L.** 2005. Homing endonuclease structure and function. Q Rev Biophys **38:**49-95.
179. **Sugisaki, H., and S. Kanazawa.** 1981. New restriction endonucleases from Flavobacterium okeanokoites (FokI) and Micrococcus luteus (MluI). Gene **16:**73-8.
180. **Summers, W.** 2005. Bacteriophage Research: Early History, p. 5-28. *In* E. Kutter and A. Sulakvelidze (ed.), Bacteriophages: Biology and Applications. CRC Press, Boca Raton.
181. **Suri, B., V. Nagaraja, and T. A. Bickle.** 1984. Bacterial DNA modification. Curr Top Microbiol Immunol **108:**1-9.
182. **Sutherland, E., L. Coe, and E. A. Raleigh.** 1992. McrBC: a multisubunit GTP-dependent restriction endonuclease. J Mol Biol **225:**327-48.
183. **Tang, M., M. Nazimiec, X. Ye, G. H. Iyer, J. Eveleigh, Y. Zheng, W. Zhou, and Y. Y. Tang.** 2001. Two forms of UvrC protein with different double-stranded DNA binding affinities. J Biol Chem **276:**3904-10.

184. **Tetart, F., C. Desplats, and H. M. Krisch.** 1998. Genome plasticity in the distal tail fiber locus of the T-even bacteriophage: recombination between conserved motifs swaps adhesin specificity. J Mol Biol **282:**543-56.
185. **Thompson, A. J., X. Yuan, W. Kudlicki, and D. L. Herrin.** 1992. Cleavage and recognition pattern of a double-strand-specific endonuclease (I-creI) encoded by the chloroplast 23S rRNA intron of Chlamydomonas reinhardtii. Gene **119:**247-51.
186. **Tiemann, B., R. Depping, E. Gineikiene, L. Kaliniene, R. Nivinskas, and W. Ruger.** 2004. ModA and ModB, two ADP-ribosyltransferases encoded by bacteriophage T4: catalytic properties and mutation analysis. J Bacteriol **186:**7262-72.
187. **Tiemann, B., R. Depping, and W. Ruger.** 1999. Overexpression, purification, and partial characterization of ADP-ribosyltransferases modA and modB of bacteriophage T4. Gene Expr **8:**187-96.
188. **Truglio, J. J., B. Rhau, D. L. Croteau, L. Wang, M. Skorvaga, E. Karakas, M. J. DellaVecchia, H. Wang, B. Van Houten, and C. Kisker.** 2005. Structural insights into the first incision reaction during nucleotide excision repair. Embo J **24:**885-94.
189. **Turmel, M., J. Boulanger, M. N. Schnare, M. W. Gray, and C. Lemieux.** 1991. Six group I introns and three internal transcribed spacers in the chloroplast large subunit ribosomal RNA gene of the green alga Chlamydomonas eugametos. J Mol Biol **218:**293-311.
190. **Waite-Rees, P. A., C. J. Keating, L. S. Moran, B. E. Slatko, L. J. Hornstra, and J. S. Benner.** 1991. Characterization and expression of the Escherichia coli Mrr restriction system. J Bacteriol **173:**5207-19.
191. **Van Roey, P., L. Meehan, J. C. Kowalski, M. Belfort, and V. Derbyshire.** 2002. Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI. Nat Struct Biol **9:**806-11.
192. **Vanamee, E. S., S. Santagata, and A. K. Aggarwal.** 2001. FokI requires two specific DNA sites for cleavage. J Mol Biol **309:**69-78.
193. **Wang, J., H. H. Kim, X. Yuan, and D. L. Herrin.** 1997. Purification, biochemical characterization and protein-DNA interactions of the I-CreI endonuclease produced in Escherichia coli. Nucleic Acids Res **25:**3767-76.
194. **Warner, H. R., P. Snustad, S. E. Jorgensen, and J. F. Koerner.** 1970. Isolation of bacteriophage T4 mutants defective in the ability to degrade host deoxyribonucleic acid. J Virol **5:**700-8.
195. **Verhoeven, E. E., M. van Kesteren, G. F. Moolenaar, R. Visse, and N. Goosen.** 2000. Catalytic sites for 3' and 5' incision of Escherichia coli nucleotide excision repair are both located in UvrC. J Biol Chem **275:**5120-3.
196. **Verhoeven, E. E., C. Wyman, G. F. Moolenaar, and N. Goosen.** 2002. The presence of two UvrB subunits in the UvrAB complex

ensures damage detection in both DNA strands. Embo J **21:**4196-205.
197. **West, D. K., L. M. Changchien, G. F. Maley, and F. Maley.** 1989. Evidence that the intron open reading frame of the phage T4 td gene encodes a specific endonuclease. J Biol Chem **264:**10343-6.
198. **Wiberg, J. S.** 1966. Mutants of bacteriophage T4 unable to cause breakdown of host DNA. Proc Natl Acad Sci U S A **55:**614-21.
199. **Vitkute, J., Z. Maneliene, M. Petrusyte, and A. Janulaitis.** 1997. BplI, a new BcgI-like restriction endonuclease, which recognizes a symmetric sequence. Nucleic Acids Res **25:**4444-6.
200. **Wyatt, G. R., and S. S. Cohen.** 1953. The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. Biochem J **55:**774-82.
201. **Wyatt, G. R., and S. S. Cohen.** 1952. A new pyrimidine base from bacteriophage nucleic acids. Nature **170:**1072-3.
202. **Yu, F., and S. Mizushima.** 1982. Roles of lipopolysaccharide and outer membrane protein OmpC of Escherichia coli K-12 in the receptor function for bacteriophage T4. J Bacteriol **151:**718-22.
203. **Zhao, L., R. P. Bonocora, D. A. Shub, and B. L. Stoddard.** 2007. The restriction fold turns to the dark side: a bacterial homing endonuclease with a PD-(D/E)-XK motif. Embo J **26:**2432-42.
204. **Zhu, Z., J. C. Samuelson, J. Zhou, A. Dore, and S. Y. Xu.** 2004. Engineering strand-specific DNA nicking enzymes from the type IIS restriction endonucleases BsaI, BsmBI, and BsmAI. J Mol Biol **337:**573-83.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 582

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-9410

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2008