Evolutionary Relationship between Different Subgroups of Restriction Endonucleases

Vera Pingoud‡ *, Elena Kubareva§, Gudrun Stengel‡, Peter Friedhoff‡, Janusz M. Bujnicki¶, Claus Urbanke#, Anna Sudina§ and Alfred Pingoud‡

‡Institut für Biochemie, Justus-Liebig-Universität, Heinrich-Buff-Ring 58, D-35392 Giessen, Germany
§A.N. Belozersky Institute of Physico-Chemical Biology and Chemistry Department, Moscow State University, Moscow 119899, Russia
¶Bioinformatics Laboratory, International Institute of Molecular and Cell Biology, ul.ks.Trojdena 4, 02-109 Warsaw, Poland
#Biophysikalisch-Biochemische Verfahren, Medizinische Hochschule Hannover, Carl-Neuberg-Str. 1, D-30623 Hannover, Germany

*Correspondence:
Dr. Vera Pingoud
Institut für Biochemie
Justus-Liebig-Universität
Heinrich-Buff-Ring 58
D-35392 Giessen
Germany

Tel.: +49 641 99 35402
Fax.: +49 641 99 35409

Email: vera.pingoud@chemie.bio.uni-giessen.de
SUMMARY

The type II restriction endonuclease SsoII shows sequence similarity with 10 other restriction endonucleases, among them the type IIE restriction endonuclease EcoRII, which requires binding to an effector site for efficient DNA cleavage, and the type IIF restriction endonuclease NgoMIV, which is active as a homotetramer and cleaves DNA with two recognition sites in a concerted reaction. We show here that SsoII is an orthodox type II enzyme, which is active as a homodimer and does not require activation by binding to an effector site. Nevertheless, it shares with EcoRII and NgoMIV a very similar DNA binding site and catalytic center as shown here by a mutational analysis, indicative of an evolutionary relationship between these three enzymes. We suggest that a similar relationship exists between other orthodox type II, type IIE and type IIF restriction endonucleases. This may explain why similarities may be more pronounced between members of different subtypes of restriction enzymes than among the members of a given subtype.
INTRODUCTION

More than 3000 different type II restriction endonucleases are known and characterized with respect to their cleavage specificities (1). This group of enzymes probably constitutes one of the largest families of enzymes with the same basic function. This makes type II restriction endonucleases (like DNA methyltransferases) ideal objects to study evolutionary relationships. Moreover, in principle, the family relationships should eventually allow predicting structural and functional features of an individual restriction endonuclease, solely on the basis of sequence comparisons. This was not obvious for a long time, because only 10 years ago, restriction endonucleases because of their little sequence conservation, were not considered to be related in evolution (2;3). But this changed due to the progress made in the analysis of sequence similarities (4-6) and, in particular, due to the increasing number of crystal structures of restriction enzymes (reviews: ref. (7;8), which made it obvious that these enzymes have very similar structural cores. Nevertheless, it has been shown recently that some genuine restriction enzymes belong to distinct nuclease superfamilies (Nuc, HNH, and GIY-YIG), which are unrelated and structurally dissimilar to each other and to the "archetypal" (P)D\_...D\_E\_X\_K superfamily (9-11). These findings make comparative studies using sequences of restriction enzymes even more challenging.

We have begun to study the restriction endonuclease SsoII recently (12-15). It is a type II enzyme (reviews: ref. (8;16), composed of identical subunits each comprising 305 amino acid residues (17). It cleaves the palindromic sequence ↓CCNGG in the presence of Mg\_2+ as indicated (18). Our interest to study this enzyme is due to its sequence similarity to EcoRII, a type IIE enzyme (review: ref. (19)), which being somewhat larger than SsoII is a dimer of identical subunits each comprising 404 amino acid residues (20;21). As a type IIE enzyme EcoRII has two DNA binding
sites, one associated with the catalytic center, the other serving as an effector site (22) (which has led to the acronym type IIE)(23-25), similar as shown also for NaeI (26-29). The recognition sequence of EcoRII is $\downarrow$CCWGG (30), which means that SsoII and EcoRII can be considered to be quasi isoschizomers. Thus, these two enzymes have a similar "genotype" (amino acid sequence) and "phenotype" (recognition sequence) which suggest a relatively close evolutionary relationship (4).

Sequence similarities between SsoII and EcoRII were first noticed by Morgan et al. (31). They had determined the sequence of PspGI, an isoschizomer of EcoRII, which is more similar in size (each subunit comprising 272 amino acid residues) and sequence to SsoII than to EcoRII. Morgan et al. identified a conserved segment of 87 amino acid residues in these three enzymes and proposed that this segment is part of a common DNA recognition domain for CCXGG sequences (31). The significance of sequence similarities between SsoII and EcoRII were emphasized by Reuter et al. (32), who had identified by a peptide scanning technique two DNA binding sites in EcoRII, one of which showed a highly significant similarity to SsoII (comprising the sequence $RR^R/KSRAGKXXE$ (DNA binding-site II)). The importance of the underlined lysine residues of the $RR^R/KSRAGKXXE$ motif for the function of EcoRII was demonstrated by a mutational analysis: the K263A/K268A and K263E/K268E variants were active in specific DNA binding, but inactive in DNA cleavage (32). In this study another DNA binding site (DNA binding site I) was identified, located in the N-terminal domain of EcoRII, not present in SsoII and PspGI, and it was proposed that this DNA binding site is associated with a potential catalytic center. The importance of this binding site for the function of EcoRII was again demonstrated by a mutational analysis: the K92A/K97A and K92E/K97E variants were severely impaired in DNA binding, but were more or less active; the E96A variant, in
contrast, was inactive in DNA cleavage. It was speculated by Reuter et al. (32), that Glu96 is part of a variant PD...D/E/XK motif, characteristic for the catalytic center of many type II restriction endonucleases (33-35), with the last lysine replaced by arginine. It is noteworthy that the sequence alignment given by Morgan et al. (31) for PspGI, SsoII and EcoRII, extends into a region that contains a potential (P)D...(S)XK...D/E motif, identified as the catalytic motif in the type IIF restriction endonucleases Cfr10I (36-38) and NgoMIV (39). The significance of this homology, however, was not realized by Morgan et al. (31). The (P)D...(S)XK...D/E sequence could represent the catalytic center not only of SsoII and PspGI, but also of EcoRII. This would imply that DNA binding site I identified by Reuter et al. (32) is the effector site and not the catalytic center.

The aim of the study presented here was to identify amino acid residues involved in DNA recognition and cleavage by SsoII. Based on sequence alignments and structural comparisons and verified by a mutational analysis we have identified two regions that are essential for the function of SsoII, and one of which is very likely to harbour the catalytic center. The sequences characterizing these regions are present in several other orthodox type II, as well as type IIE enzymes (which require binding of a substrate at the active site and the effector site) and IIF enzymes (which require binding of a substrate at two active sites formed by four identical subunits), that recognize CCGG, CCNGG or CCWGG in different sequence contexts, an observation that has implications for the evolution of these subtypes of type II restriction endonucleases.

EXPERIMENTAL PROCEDURES

*Site-directed Mutagenesis of SsoII Variants*
Site-directed mutagenesis of the SsoII gene was performed by a PCR-based technique (40). The mutant genes were sequenced and found to contain only the mutation desired.
Protein Expression and Purification

The SsoII overproducing strain was kindly provided by Dr. A. Karyagina (Institute of Agricultural Biotechnology, Moscow). SsoII and its variants were expressed in Escherichia coli JM109 and purified to homogeneity as described by Sheflyan et al. (14).

Analytical Ultracentrifugation Experiments

Analytical ultracentrifugation was done in a Beckman XL-A Analytical Ultracentrifuge equipped with a photoelectric scanner. All experiments were done at 4°C in 10 mM Tris/HCl, pH 8.0, 50 mM NaCl.

Sedimentation velocity experiments were done in an 8-place An50 Ti rotor with double sector centerpieces made of charcoal filled epon at 45000 rpm.

To evaluate binding of double-stranded d(GCTGCCAACCTGGGTCTAAC) (US/LS-20a) to SsoII the concentrations of free and bound oligo US/LS-20a were extracted from the measured profiles A(x,t) by fitting the equation

\[
A_i(x_m,0) = A_{i_0} + A_{i_1} \exp(-s_i x_m / \sigma_k) \sin(\omega x_m) + A_{i_2} \cos(\omega x_m)
\]

with \(x_m\) : position of the meniscus; \(A_{i_0}\) : absorption of the i-th sedimenting boundary at the beginning of the experiment; \(s_i\) : sedimentation coefficient of the i-th sedimenting boundary; \(\sigma_k\) : a parameter describing the width of each measured sedimentation profile; \(\omega\) : angular velocity of the rotor. This fit corresponds to a description of the sedimenting boundaries by a gaussian error function and yields the absorption for free oligo US/LS-20a and the sum of the absorptions for total SsoII and bound oligo US/LS-20a at the beginning of the sedimentation run.

For sedimentation equilibrium analysis samples of 120 µl were run in 6-channel centerpieces
(column height approximately 3 mm) at the specified speed until no change in concentration distribution could be observed for at least 12 hours. Scans from these 12 hours were averaged to reduce random noise.

A simple aggregation model will describe the association of \( n \) dimers of \( Sso\text{II} \) to form a single aggregate. For such a reaction the law of mass action can be written as:

\[
A(ted) = \sum_{i=1}^{2} A_{i}(0) \cdot \left( \frac{\sigma_k \sqrt{2}}{\sigma_k^2} \right)^{\frac{1}{2}} \int_{-\infty}^{\infty} \left( \frac{1}{\sqrt{2\pi} \sigma_k} \right)^2 e^{-\frac{(x-m)^2}{2\sigma_k^2}} \, dx
\]

with \( Sso\text{II}_n \) being the \( n \)-mer aggregate and \( K_{agg} \) the bimolecular equilibrium constant for aggregation.

For evaluation of the sedimentation equilibrium the distribution of each species is calculated by numerically solving

\[
K_{agg}^{-1} = \frac{[RSso\text{II}_n]}{[RSso\text{II}]}
\]

\[
c_i(x) = \exp \left( \frac{M_i(1-\rho)}{RT} \omega^2 \right)
\]

: radial position (of the middle of the cell), \( c_i(x) \): partial specific volume

where all \( c_i(x) \) must obey to the above law of mass action. The partial specific volume for \( Sso\text{II} \) was calculated from the amino acid composition to be \( 7.25 \times 10^{-4} \) m\(^3\)/kg.

**Gelfiltration**

Gelfiltration experiments were performed at room temperature on a Merck-Hitachi system using a Superdex 75 column (Pharmacia), equilibrated with 25 mM Tris/HCl, pH 8.0, 100 mM NaCl, 10 mM CaCl\(_2\). Free \( Sso\text{II} \) or \( Sso\text{II} \) in complex with oligo US/LS-20a were diluted in the same buffer and loaded onto the column. Chromatography was carried out at a flow-rate of 1 ml/min at 20 °C. Elution was monitored by absorbance at 280 nm. The molar masses were determined by interpolation, using a calibration curve of proteins of known molar mass.
(Molecular weight marker kit, Sigma).
Circular Dichroism Spectroscopy

Circular dichroism spectra were recorded in 10 mM Tris, 50 mM NaCl in a Jasco J-710 dichrograph at 20 °C in a cylindrical cuvette of 0.05 cm path length.

DNA Cleavage Assay

The DNA cleavage activity of wild type SsoII and the mutants were determined by performing cleavage experiments in cleavage buffer (10 mM Tris/HCl, pH 7.5, 50 mM NaCl, 1 mM DTT, 10 mM MgCl₂, 0.1 mg/ml bovine serum albumin) at 37°C with ³²P-labeled PCR products as substrate.

Cleavage experiments designed to compare the rates of cleavage of one- and two-site substrates by wild type SsoII were performed with two different PCR products containing one recognition sequence each with a different flanking base pair (GCTGGG (325 bp) and CCGTGGG (221 bp)) and one PCR product containing both recognition sites (442 bp). Cleavage experiments which were carried out to determine the activity of the SsoII variants were carried out with the 325 bp substrate. After incubation for a specified time, reaction products were analyzed by electrophoresis on 10% polyacrylamide gels. The gels were subsequently subjected to autoradiography using an instant imager (Canberra Packard).

Electrophoretic Mobility Shift Assay

For electrophoretic mobility shift assays SsoII and the respective variants were mixed with binding buffer (10 mM Tris/HCl, pH 8.5, 100 mM NaCl, 5 mM CaCl₂, 0.1 mg/ml bovine serum albumin, 10% (v/v) glycerol) in a total volume of 10 µl containing in addition 0.1 µg of poly(dI-dC) (Pharmacia), and ³²P-labeled double-stranded d(GCTGCCACCTGGGTCTAAAC), oligo US/LS-20b. After incubation for 30 min at 20 °C the samples were subjected to electrophoresis on 12% polyacrylamide gel. After electrophoresis for five hours at 80 V (10 V/cm) in 20 mM
Tris/acetate, pH 8.5, 5 mM CaCl₂, gels were subjected to autoradiography using an instant imager (Canberra Packard).

**Sequence Analysis and Structure Prediction**

The multiple alignment was obtained based on PSI-BLAST searches of the non-redundant sequence database at the NCBI (http://www.ncbi.nlm.nih.gov) and edited manually based on consensus secondary structure prediction and threading results reported by the Meta Server (41;42) with the links to the individual methods provided therein and at the website http://bioinfo.pl/meta/. The alignment of NgoMIV and Cfr10I was based on the structural superposition of the atomic coordinates. The neighbour-joining tree (43) was estimated from 100 bootstrap resamplings of the conserved alignment.

Homology modeling of the core region (residues 115-257) of SsoII was performed using Swiss-PdbViewer and Swiss-model (44) based on the coordinates of the NgoMIV-product complex (39). A series of alternative models of the SsoII monomer were submitted to the Verify3D server (45) to evaluate compatibility of the residues with the environment. The best structure was used to model the interactions of SsoII with the cleaved target DNA. At the outset, the coordinates of the double-stranded d(GCGCCGGCGC) containing the NgoMIV target sequence were copied to the model from the template structure. Subsequently, the SsoII-DNA complex was duplicated and positioned in such a way, that one of the subunits remained superimposed onto the corresponding NgoMIV monomer, while the other was shifted and rotated to reflect insertion of one base pair in the middle of the DNA sequence. It was immediately recognized that preservation of the NgoMIV-like structure of the DNA and orientation of the active sites in respect to the “CC” element in both SsoII monomers lead to unresolvable steric clashes and loss of contact of Arg186 and Arg188 side chains with the “GG” element. Thus, the 185-189 loop of SsoII was separately superimposed onto the 190-
195 loop of NgoMIV to mimick its interaction with the “GG” element. Subsequently, the two subunits were mutually rotated together with the DNA to reduce the overlap at the dimer interface and to bring the ends of the displaced 185-189 loop close enough to the original attachment points to allow reconnection of the polypeptide backbone without distortion of the neighboring regions. As an outcome, the model of the SsoII-DNA complex contains two regions of interaction with the “CC” and the “GG” elements that closely resemble those of the NgoMIV template structure, but which could not be simultaneously superimposed onto it.

The final model of the monomer was refined using GROMOS (46), duplicated, and docked onto the DNA created from two NgoMIV half-sites “GCGCC” with an additional AT base pair manually inserted in the middle.

RESULTS

We show here that SsoII shares sequence similarities with EcoRII, Cfr10I and NgoMIV (among others). The most interesting aspect of this is that EcoRII, Cfr10I and NgoMIV require simultaneous binding of two substrates (or one substrate with two recognition sites) for efficient cleavage. Whereas EcoRII is a type IIE enzyme, which are homodimeric proteins with two distinct binding sites for their respective recognition sequences, one being the active site, the other the effector site, Cfr10I and NgoMIV are type IIF enzymes, which are composed of four identical subunits that form two active sites which in a coordinate manner cleave the two recognition sites (review: (47)). In order to characterize the relationship between SsoII and the type IIE (EcoRII) and type IIF (Cfr10I and NgoMIV) enzymes, it had to be investigated to what subtype of type II restriction endonucleases SsoII belongs and to find out whether the sequence similarities between SsoII on one hand and type IIE and IIF enzymes on the other hand concern functionally important residues.
Analytical Ultracentrifugation and Gelfiltration Experiments-Orthodox type II restriction enzymes are homodimeric enzymes. The quaternary structure of SsoII is not known. In order to determine the native molar mass of SsoII, analytical ultracentrifugation and gelfiltration experiments were carried out.

Sedimentation velocity runs with free SsoII showed that most of the protein (> 80%) sedimented with 5 S. Since the largest sedimentation coefficient for a protein with an Mr of 35.9 kDa is 3.77 S, the value of 5 S shows that the predominant form of SsoII is a dimer. A minor component (< 20%) sedimented with approx. 20 S indicating the presence of larger aggregates (data not shown).

Fig. 1A shows the result of a sedimentation equilibrium experiment with SsoII. Assuming an aggregation equilibrium, the concentration profile can best be described by a reaction of dimers, the predominant species at µM concentrations, interacting in a concentration dependent manner with each other to form an aggregate consisting of approx. 14 dimers, with a bimolecular equilibrium constant, $K_{agg}$, of 1 x 10^5 M^-1 which would yield half of the dimer being engaged in aggregates at a total concentration of 3 x 10^-4 M. The presence of significant amounts of monomers is incompatible with the measured protein concentration profile as is an association of dimers forming only tetramers.

Fig. 1B shows the sedimentation profiles of a mixture of SsoII and excess oligo US/LS-20a, a double-stranded eicosadeoxynucleotide containing one recognition site in a central position. From the evaluation of the absorption of the two sedimenting boundaries a stoichiometry of one oligo US/LS-20a binding to one SsoII dimer can be deduced.

Gelfiltration on a Superdex 75 column showed that SsoII co-elutes with bovine serum albumin (Mr = 66 kDa) indicating that the quaternary structure of SsoII is that of a dimer (Mr
= 72 kDa). In the presence of Ca\(^{2+}\) and an excess of oligo US/LS-20a, the SsoII x DNA complex eluted with an apparent M\(_r\) of 80 kDa, which is in good agreement with the calculated M\(_r\) of 84 kDa for the 1:1 complex (data not shown).

The results of the analytical ultracentrifugation and the gel filtration experiments demonstrate that SsoII like orthodox type II restriction endonucleases is a homodimer, both in the absence and presence of DNA.

Our finding that SsoII is a homodimer is somewhat at variance with a result obtained for a very close relative of SsoII, namely Ecl18kI which differs only in one position (pos. 232 where SsoII has an isoleucine and Ecl18kI a valine residue) (48): based on a gelfiltration analysis Ecl18kI was identified to be a homotetramer (49). Of course, it cannot be excluded that a valine for isoleucine substitution could change a dimer to tetramer equilibrium.

**Steady-state DNA Cleavage Experiments**

SsoII shows sequence similarity to EcoRII, a typical type IIE restriction endonuclease, which requires binding to two recognition sites for efficient cleavage (22). To find out whether SsoII behaves like a type IIE enzyme, DNA cleavage experiments were carried out with substrates containing one or two sites. PCR products containing one (221-mer, 325-mer) or two sites (442-mer) are cleaved with rates dependent on the sequence context, CCC\text{TGG} (k\(_{app}\) = 0.26 ± 0.03 min\(^{-1}\)) is cleaved faster than GCC\text{TGG} (k\(_{app}\) = 0.15 ± 0.02 min\(^{-1}\)), but independent of whether one or two sites are present (Fig. 2). Cleavage of the 325-mer (which contains only one site) could not be stimulated in trans, as observed for EcoRII (50), by addition of oligo US/LS-20a in concentrations varied from 0.5 to 5-fold excess; in contrast, with increasing concentration of the oligodeoxynucleotide an inhibition of the cleavage of the 325-mer was observed (data not shown). The cleavage of the
1-site and 2-site substrates was investigated at different ionic strengths. With increasing ionic strength (from 50 mM to 150 mM NaCl) SsoII cleaves the substrates with one (221 bp, CCCTGGG) or two recognition sites (442 bp, CCCTGGG and GCCTGGG) with reduced but equal rates (data not shown). These results confirm that SsoII, in spite of its sequence similarity to EcoRII, is an orthodox type II enzyme.

Alignment of the sequence of SsoII with the sequences of other type II restriction endonucleases - Sequence similarities between SsoII (recognition sequence: ↓CCNGG, PspGI (↓CCWGG) and EcoRII (↓CCWGG)) had been identified over a stretch of approximately 90 amino acid residues (amino acid residues 108 to 196 in SsoII) by Morgan et al. (31) leading to speculations that this segment is part of a common DNA recognition domain for CCXGG, and should also be found in the amino acid sequence of BstNI (CC↓WGG), MvaI (CC↓WGG) and ScrFI (CC↓NGG). Our inspection of the sequences of ScrFI (51) and of MvaI (A. Karyagina, pers. comm.) showed that there is no significant similarity between these sequences and the sequences of SsoII, PspGI, and EcoRII; the sequence of BstNI is not yet available. On the other hand, it was recently observed that eight different type II restriction enzymes (including SsoII, PspGI, EcoRII and among others NgoMIV, whose cocrystal structure is known), which all have recognition sites with adjacent guanine residues, harbour a KX3RXXRX6E motif (8). The cocrystal structure of NgoMIV indeed shows that the Arg191 and Arg194 residues recognize the guanines (underlined) of the GCCG GCC recognition sequence (39). We have now systematically analyzed all available sequences of type II restriction enzymes, including sequences of presumptive type II restriction enzymes obtained in the course of genome sequencing projects, for sequence similarity to SsoII. This analysis came up with 11 bona fide type II restriction enzymes (AccIII, BsrFI, Cfr10I, EcoRII, HpyAo263, Kpn2I, MjaVIP, NgoMIV, PspGI, SgrAI and SsoII) and three presumptive type II restriction enzymes (CteIP, DvuIP,
*Nme* Ao1499) with clustered regions of significant sequence similarity (Fig. 3). It was intriguing to note that the recognition sequences of these enzymes all contain CCGG or CCNGG or CCWGG, and, furthermore, that these enzymes belong to different subtypes, viz. orthodox type II enzymes (like *Sso*II), type IIE enzymes (like *Eco*RII) and type IIF enzymes (like *Cfr*10I and *Ngo*MIV).

We failed to detect a typical (P)D...D/E XK motif, which characterizes the active site of many type II restriction endonucleases (6;33-35), in the sequence of any of the endonucleases represented in Fig. 3. However, we found a very well conserved (P)D...(S)XK...D/E motif, which has been shown first for *Cfr*10I to be a variant of the (P)D...D/E XK motif (37) and later identified as such in the *Ngo*MIV cocrystal structure (39). The (P)D...(S)XK...D/E motif is 100% conserved in all the sequences shown in the alignment, giving credence to the supposition that it represents the active center not only in *Ngo*MIV and *Cfr*10I but also in the other enzymes. Between the (presumptive) catalytic residues K (Lys187 in *Ngo*MIV) and D/E (Glu201 in *Ngo*MIV) there are three also 100% conserved R/K (Arg191 in *Ngo*MIV), D/E (Asp193 in *Ngo*MIV) and R (Arg194 in *Ngo*MIV) residues, which in *Ngo*MIV are involved in hydrogen bonds to the inner cytosine and guanine residues (underlined) of the recognition sequence G↓CCGGC. It is tempting to conclude that these conserved R/K, D/E and R residues have a similar function in the other enzymes shown in Fig. 3, as their counterparts in *Ngo*MIV.

Another region with almost 100% conservation in the alignment of Fig. 3 is around Glu70 in the *Ngo*MIV sequence. In the cocrystal structure this residue is located close to the active center of this enzyme. Nearby is Gln63 which is involved in a hydrogen bond to the last base
of the recognition sequence $\text{G↓CCGGC}$ \cite{39}. The equivalent for this residue in $\text{SsoII}$, $\text{PspGI}$ and $\text{EcoRII}$ is a serine residue, located in the middle of a sequence conserved among these three enzymes: $\text{RXS\text{RaGKXXE}}$. This sequence was identified by Reuter et al. \cite{32} to be involved in DNA binding by $\text{EcoRII}$ (DNA binding site II); it is tempting to speculate that this is also true for the other enzymes whose sequence is shown in the alignment of Fig. 3.

**Mutational analysis of amino acid residues in DNA binding and cleavage**—Based on the multiple sequence alignment shown in Fig. 3, Arg116, Arg117, Arg119 and Lys122 as well as Arg186 and Arg188 are candidates for amino acid residues involved in DNA binding and recognition by $\text{SsoII}$, whereas Glu125, Asp160, Lys182, Glu191 and/or Glu195 could be responsible for DNA cleavage. These residues were subjected to a mutational analysis. In addition, Glu187, Glu233 and Lys235 were included in this analysis: Glu187, because it is also a well conserved amino acid residue (Fig. 3), Glu233 and Lys235 as alternative candidate amino acid residues to be involved in catalysis (these residues are part of a variant PD...\text{D}/\text{E}XK motif: $\text{P}^{194}\text{E}...^{233}\text{E}N^{235}\text{K}$).

The variants carrying an alanine substitution in place of the wild type amino acid residue were produced by site-directed mutagenesis, purified to homogeneity and analyzed with respect to their DNA binding and cleavage activity.

DNA binding of the variants was analyzed by gel electrophoretic mobility shift assays using oligo US/LS-20b in the presence of $\text{Ca}^{2+}$. Under these conditions $\text{SsoII}$ binds specifically to DNA containing a recognition site, whereas in the absence of $\text{Ca}^{2+}$ there is no specific binding, as multiple band shifts are observed with a PCR fragment (G. Stengel, unpublished), similarly as reported for $\text{EcoRV}$ \cite{52}. Fig. 4 shows the results of the gel shift experiments with oligo US/LS-20b: the variants R116A, R117A, R119A, R186A and R188A show no or little,
the K122A, E125A, D160A and E187A variants reduced DNA binding, whereas the K182A, E191A, E195A, E233A and K235A display a very similar affinity for DNA as the wild type enzyme (Fig. 4 and Table I), which was determined in several titrations to be 40 ± 10 nM (data not shown). The inability of some variants to bind to DNA specifically is not due to an aberrant folding of the variants, because it was demonstrated that these variants have an identical circular dichroism spectrum as the wild type enzyme (data not shown).

DNA cleavage by the variants was analyzed using a 325 bp PCR product with one SsoII recognition site (Fig. 5). As expected, the variants defective in DNA binding in the presence of Ca²⁺, namely R116A, R117A, R119A, R186A and R188A, are completely inactive in DNA cleavage. In addition, E125A, D160A, K182A, E187A and E195A turned out to be unable of cleaving DNA, although they can interact with DNA in a specific manner. Detailed kinetic experiments carried out over more extended incubation times demonstrated a residual DNA cleavage activity only for the E195A variant in the range of 2 % of the wild type cleavage activity (Fig. 6 and Table I). The small but nevertheless detectable activity of the E195A variant suggested to us that Glu191 could partially "rescue" this mutant, in particular as in the alignment shown in Fig. 3 two restriction enzymes do not have a glutamic acid residue corresponding to Glu195 but instead only one corresponding to Glu191. We, therefore, produced the double mutant E191Q/E195Q which turned out to be completely inactive (V. Pingoud, unpublished). It is important to note that E191A, E233A and K235A show largely unaltered DNA binding and cleavage activity. This result could be interpreted to mean that Glu125, Asp160, Lys182 and Glu195 have a direct catalytic function, not however, Glu191, Glu233 and Lys235. The low activity of the K122A variant could be due to the fact that Lys122 is located close to the cluster of Arg residues (Arg116 to Arg119) which are essential for DNA binding and cleavage.
The results of our mutational analysis of the active site residues of SsoII are in very good agreement with results obtained for StyD4I, a very close relative of SsoII, which differs only in 15 positions from SsoII, 13 of which are due to an N-terminal extension of StyD4I (53): Siksnys and colleagues have shown that variants of StyD4I corresponding to the SsoII variants D160A, K182A, R186A, R188A and E195A behave very similarly to the SsoII variants (V. Siksnys, pers. comm.).

DISCUSSION

For a long time type II restriction endonucleases were considered to be unrelated in evolution, with the exception of some obviously homologous isoschizomers. Only after the determination of the crystal structure of EcoRV (54) and the comparison with the crystal structure of EcoRI (55;56) did it become clear that there is a structural similarity between different seemingly unrelated enzymes. Due to divergent evolution, different subtypes exist with differences in quaternary structure, domain organization and active site architecture on one side and mechanistic details on the other side.

The majority of type II restriction enzymes are homodimers which cleave the DNA within or immediately adjacent to their recognition site and do not depend on binding to a second copy of their recognition site for maximum activity. However, some of them comprise four identical subunits. Like SfiI (the first type IIF enzyme discovered), Cfr10I and NgoMIV cooperate in the concerted cleavage of two recognition sites (57-59). Others depend on an effector site which must be occupied for efficient DNA cleavage (type IIE, like NaeI and EcoRII); the DNA in the effector site, however, is not cleaved simultaneously with the DNA in the cleavage site (59).

When we started to study SsoII it was not clear, whether it is an orthodox type II restriction
endonuclease or not, but we soon noticed that it shares sequence similarity with bona fide type IIE (EcoRII) and type IIF (Cfr10I and NgoMIV) enzymes in functionally relevant regions (EcoRII: (32), Cfr10I: (36-38), NgoMIV: (39)), as shown in Fig. 3. Our analysis of the quaternary structure of SsoII by analytical ultracentrifugation and gel filtration presented here, demonstrates that it is a homodimer, in the absence as well as in the presence of DNA. In this respect it behaves like an orthodox type II restriction endonuclease or like a type IIE enzyme, not however, like a type IIF enzyme. The fact that SsoII, unlike EcoRII (22;60;61) is not stimulated by a second site on a DNA substrate, as shown here, excludes that it is a type IIE enzyme, but rather an orthodox type II restriction endonuclease.

The active site of SsoII and EcoRII has not yet been identified so far. The inspection of the SsoII amino acid sequence yielded two candidate motifs for the active site; (1) (P)\(^{194}\)E...\(^{233}\)EN\(^{235}\)K and (2) (P)\(^{160}\)D...(S)A\(^{182}\)K...\(^{195}\)E. No plausible candidates for motifs characteristic for other nuclease families were detected. The first motif is unusual in as much as it has a glutamic acid residue where so far only an aspartic acid residue has been found. The mutational analysis indeed shows that neither Glu233 nor Lys235 are essential for DNA cleavage, excluding that this sequence represents the catalytic center. The second motif has been identified so far only as a variant (P)D...\(^{D/E}\)XK motif in the type IIF restriction endonucleases Cfr10I (36-38) and NgoMIV (39). The alignment in Fig. 3, however, demonstrates that this variant (P)D...\(^{D/E}\)XK motif is embedded in a stretch of similar amino acid residues common to 11 restriction enzymes (and three putative restriction enzymes) which all have CCNGG, CCWGG or CCGG in their respective recognition sequences, among them EcoRII. The mutational analysis of Asp160, Lys182 and Glu195 in SsoII demonstrates that this
variant (P)D...D/E/XK motif most likely is the true catalytic motif, because Asp160, Lys182 and Glu195 were found in the present study to be essential for cleavage.

It can be expected that the sequence homologies apparent in the sequence alignment (Fig. 3) reflect structural homologies, which should manifest themselves in a secondary structure similar to that of NgoMIV observed in the cocrystal structure of NgoMIV (39). Indeed, a secondary structure prediction carried out for SsoII and, for comparison, also for EcoRII, shows that in the region of interest (SsoII: amino acids 123 to 232; EcoRII: amino acids 270 to 370), there is a close correspondence of secondary structure elements between SsoII, EcoRII and NgoMIV (as deduced from the cocrystal structure). As shown in Fig. 8, the structural core, which is characteristic for many type II restriction endonucleases (7;28;62), is likely to be also present in SsoII and EcoRII. Projected onto this structural core, the presumptive functionally important amino acid residues are found in similar locations as observed in NgoMIV, if one takes into account that the borders of individual secondary structure elements are not easily predicted.

The quality of the multiple sequence alignment and of the secondary structure prediction suggests that it should be possible to model the core structural elements of SsoII using the NgoMIV structure as a template (39). For that purpose, the SsoII sequence was manually threaded onto the NgoMIV structure using multiple sequence alignment and secondary structure prediction to superimpose conserved and therefore presumably functionally important amino acid residues of the target and the template. The final model of an SsoII monomer comprising amino acid residues 115 to 257 was obtained after several cycles of modelling and energy minimization. For the modelling of the complex, the DNA of the NgoMIV-DNA complex was docked onto the modelled SsoII monomer, the DNA expanded by a central A:T base pair
(to convert the *NgoMIV* recognition site, GCCGGC, into an *SsoII* recognition site, CCNGG) and a second *SsoII* monomer positioned on the DNA in a symmetrical fashion. No attempts were made to refine the dimer interface, the protein-DNA interface and the structure of the DNA beyond the relief of the most severe steric clashes, since we do not believe that this is justified given the relatively low degree of sequence similarity between *SsoII* and *NgoMIV* and the methodology available. On the other hand we trust that the model is sufficiently detailed to illustrate that the sequence alignment and the secondary structure prediction can be translated into a structural model which allows us to visualize how *SsoII* might interact with DNA (Fig. 8). According to this model Arg186, Glu187 and Arg188 as their counterparts in *NgoMIV* are involved in base recognition, whereas Arg116, Arg117 and Arg119 seem to be in an ideal position to interact with the region at the end of the recognition sequence.

**CONCLUSIONS**

Our results allow concluding that *SsoII*, an orthodox type II restriction enzyme, has a similar DNA binding site and catalytic center as the typical type IIF enzyme *NgoMIV*. The pronounced sequence similarities between *SsoII* and *EcoRII*, a typical IIE enzyme, suggest that one can extend this conclusion also to this enzyme. This would imply that Arg262, Lys263, Arg265, Lys268, Lys328 and Arg330 of *EcoRII* are involved in DNA recognition, whereas Asp299, Lys324 and Glu337 are responsible for DNA cleavage, and furthermore, that DNA binding site I identified by Reuter *et al.* (32) is the allosteric effector site.

An important outcome of our study concerns the evolutionary relationship of *SsoII*, *EcoRII* and *NgoMIV* as representatives of three subfamilies within the family of type II restriction endonucleases. We suggest that some members of these subfamilies share a not too distant common ancestor (Fig. 9), which presumably was a homodimeric enzyme (similar to *SsoII*).
By acquisition of an extra domain (as it is seen in \textit{NaeI} (29)) a type IIE enzyme like \textit{EcoRII} was derived. Alternatively, by formation of a new interface for protein-protein interaction tetramerization was made possible giving rise to representatives of type IIF enzymes like \textit{Cfr10I} and \textit{NgoMIV}. We believe that this scenario happened many times starting from a homodimeric restriction enzyme, which led to different groups of type IIE and IIF enzymes (63;64). This explains why sequence similarities are sometimes more pronounced between members of different subtypes of restriction enzymes than among the members of a subtype. In this respect, \textit{SsoII} evolutionarily is more related to \textit{EcoRII} and \textit{Cfr10I} or \textit{NgoMIV}, than \textit{EcoRII} is to \textit{NaeI} (another type IIE enzyme).

\textit{Acknowledgements}-We thank Dr. A. Karyagina for supplying the \textit{SsoII} overproducing strain and making available the sequence information of \textit{Mvac}, Mrs. C. Conzelmann for preparation of some of the variants, Dr. M. Reuter, Mrs. M. Mücke and Dr. V. Siksnys for communicating unpublished results, and Dr. A. Jeltsch for critical reading of the manuscript. The technical assistance of Nadine Thome is gratefully acknowledged. This work was supported by the Deutsche Forschungsgemeinschaft (Pi 122/13-3), the Fonds der Chemischen Industrie and the Deutscher Akademischer Austauschdienst.
FIGURE LEGENDS

FIG. 1. **Analytical ultracentrifugation.** (A) Sedimentation equilibrium of 6 µM SsoII at 10000 rpm (u). The smooth line represents a theoretical concentration profile and (s) the corresponding residuals for a SsoII dimer aggregating to form a 14-mer aggregate with an equilibrium constant of $1 \times 10^5$ M$^{-1}$. Note the different scale for the residuals. (B) Sedimentation velocity experiment with 2.5 µM SsoII dimer and 6.5 µM double-stranded d(GCTGCCAACCTGGGTCTAAC). Scans were taken at an interval of 480 s. The smooth lines represent theoretical concentration distributions calculated as a sum of two gaussians where the fast moving boundary contains 2.5 µM SsoII dimer and 2.1 µM double-stranded eicosadeoxynucleotide indicating a 1:1 stoichiometry.

FIG. 2. **Cleavage of radioactively labelled substrates with one or two recognition sites by SsoII.** (A) 100 nM PCR products containing either one (221 bp, CCCTGGG, left; 325 bp, GCCTGGG-, right) or two recognition sites (442 bp, CCCTGGG and GCCTGGG, middle) were incubated with 25 nM SsoII. Aliquots were withdrawn after 0, 2.5, 5, 7.5,10,15 and 20 min., respectively, and analyzed by electrophoresis, followed by autoradiography using an instant imager. (B) Disappearance of the two 1-site and the 2-site substrates was evaluated quantitatively for the two 1-site substrates (221 bp, 1, 325 bp, n) and the 2-site substrate (442 bp, ). It is evident that the 2-site substrate (442 bp) is cleaved as fast as the 1-site substrate with the CCCTGGG sequence (221 bp).

FIG. 3. **Alignment of partial amino acid sequences of SsoII and 13 other bona fide (recognition sequence indicated) or presumptive type II restriction endonucleases.** Similar or identical amino acid residues in the 14 sequences are indicated in gray or black. Amino acid residues which were subjected to a mutational analysis in SsoII are indicated on top of the figure in black (presumably involved in catalysis) or gray (presumably involved in DNA
binding); their position in the SsoII sequence is indicated on top. The alignment is subdivided in two groups: the top group comprises restriction enzymes that recognize CCNGG or CCWG, the bottom group CCGG.

FIG. 4. Electrophoretic mobility shift analysis of SsoII and its variants. 50 nM SsoII or its variants were incubated with 100 nM 32P-labeled double-stranded d(GCTGCCACCTGGGTCTAAC). DNA bound and unbound SsoII species were separated by electrophoresis and subsequently analyzed using an instant imager.

FIG. 5. Cleavage of a 325 bp PCR product by SsoII and its variants. 100 nM 32P-labeled 325 bp PCR product containing one recognition site was incubated with 25 nM SsoII or its variants for 15 min at 37°C. Cleavage products were separated by electrophoresis and analyzed using an instant imager.

FIG. 6. Cleavage of a 325 bp PCR product by the SsoII variant E195A. 100 nM 32P-labeled 325 bp PCR product containing one recognition site was incubated with 25 nM SsoII variant E195A. Aliquots were withdrawn at the time indicated and analyzed by electrophoresis. Quantitative evaluation was performed using an instant imager.

FIG. 7. Secondary structure prediction for SsoII and EcoRII. Shown on top is the secondary structure of NgoMIV, as determined by crystallography of the NgoMIV-product complex. Indicated in gray are secondary structure elements. Residues indicated by circles or crosses are involved in DNA binding or catalysis, respectively (demonstrated for NgoMIV by the cocrystal structure, suggested for SsoII and EcoRII by a mutational analysis and/or by analogy to NgoMIV). Residue no. 61 in SsoII is known to be involved in DNA binding based on a crosslink (15), residues 261-263/265 in EcoRII are in DNA binding site I (32).

FIG. 8. Comparison of the NgoMIV cocrystal structure with an SsoII model. Shown is a
detail of the structure to illustrate the secondary structure elements (yellow) and amino acid residues involved in DNA recognition (R116, R117, R119, K122, R186, E187, R188) and cleavage (E125, D160, K182, E191, E195). In the SsoII model the position of Arg116, 117 and 119 are indicated by balls, to demonstrate that their side-chain positions cannot be predicted reliably. The underlined cytosine and guanine residues of the recognition sequence (GCGGCG for NgoMIV, CCGG for SsoII) are highlighted.

**FIG. 9. Evolutionary tree for SsoII and its homologues.** The bar at the bottom represents the evolutionary distance scale. The numbers at the nodes indicate % bootstrap support of the individual branching points. Evolutionary events resulting in developing novel features by the type IIE (EcoRII) and type IIF (NgoMIV) branches are indicated. Although the presented tree is unrooted, the position of the root is inferred to be between the SsoII/EcoRII branch and the type IIF (NgoMIV) branch based on the analysis using sequences of enzymes cleaving sequences different than CCXGG or CCGG as an outgroup (6; J.M.Bujnicki, unpublished data).
Table I

*Relative DNA cleavage and binding activity of SsoII variants with alanine substitutions at positions of presumptive functional importance*

Cleavage activity was determined by performing cleavage experiments with 100 nM 32P-labeled 325 bp PCR product containing one recognition site (GCC\textsuperscript{TGG}G) and 25 nM wt or mutant SsoII at 37°C for 60 min. \(k_{\text{app}}\)-values were calculated by measuring the initial velocity of cleavage. A value of 0.15 ± 0.02 min\(^{-1}\) was obtained for the wt enzyme.

Binding activity was determined using electrophoretic mobility shift assays by incubating 100 nM 32P-labeled double-stranded d(GCTGCCAC\textsuperscript{CCTGG}GTCTAAC) with increasing concentrations of SsoII or its variants. The binding isotherms were evaluated in terms of \(K_d\)-values.

<table>
<thead>
<tr>
<th>Variant</th>
<th>Cleavage Activity [%]</th>
<th>Binding Activity [(K_d), nM]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wt</td>
<td>100</td>
<td>80</td>
</tr>
<tr>
<td>R116A</td>
<td>0</td>
<td>n.d.b.</td>
</tr>
<tr>
<td>R117A</td>
<td>0</td>
<td>n.d.b.</td>
</tr>
<tr>
<td>R119A</td>
<td>0</td>
<td>n.d.b.</td>
</tr>
<tr>
<td>K122A</td>
<td>10</td>
<td>160</td>
</tr>
<tr>
<td>E125A</td>
<td>0</td>
<td>140</td>
</tr>
<tr>
<td>D160A</td>
<td>0</td>
<td>300</td>
</tr>
<tr>
<td>K182A</td>
<td>0</td>
<td>60</td>
</tr>
<tr>
<td>R186A</td>
<td>0</td>
<td>800</td>
</tr>
<tr>
<td>E187A</td>
<td>0</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>R188A</td>
<td>0</td>
<td>500</td>
</tr>
<tr>
<td>E191A</td>
<td>100</td>
<td>90</td>
</tr>
<tr>
<td>E195A</td>
<td>2</td>
<td>60</td>
</tr>
<tr>
<td>E233A</td>
<td>100</td>
<td>90</td>
</tr>
<tr>
<td>K235A</td>
<td>75</td>
<td>80</td>
</tr>
</tbody>
</table>

n.d.b., no detectable binding
REFERENCES


46. Scott, W. R. P., Hunenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennen, J.,


Pingoud, V., et al., Figure 2

A  
<table>
<thead>
<tr>
<th>1 site</th>
<th>2 sites</th>
<th>1 site</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCCTGGG</td>
<td>GCCTGGG</td>
<td>GCCTGGG</td>
</tr>
</tbody>
</table>

- 221 bp
- 169 bp
- 52 bp
- 1 site
- 442 bp
- 325 bp
- 184 bp
- 141 bp

B  

% Product vs. Time [min]

0 5 10 15 20
Pingoud et al., Figure 3

DNA binding/Catalysis

<p>| | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RR</td>
<td>R</td>
<td>E</td>
<td>D</td>
<td>K</td>
<td>R</td>
<td>E</td>
<td>R</td>
<td>E</td>
<td>R</td>
<td>E</td>
<td>R</td>
<td></td>
</tr>
</tbody>
</table>

SsoII  CCGGG : 111-SNTG SRRKNNEGLNEELAG1MGAGL-(16)-QIGKLVWVHAGV-(6)-KRNTMILKQTVLILRQEPRQPVW1NGG

PspGI  CCGGG : -91-SFQAA FSQQSNEEFTKLNKGG-(14)-TEGEKDPIEVSV-(6)-PSSAIITTVKVRKREDGEQIQIIR

EcoRII  CCGGG : 257-SVSA SRSRSLMEEHFLERG-9-GNKKPQPSA-(9)-VYNKLAKAQTLRDKPQERQER

NmeHolo499 : 245-SFTK SARKPPLSKKSKFTG-8-LGKSMNQGSS-15-TQELGSLNCKNCKQK

DvuIP  CCGGG : -7-SVLP SRRKNNLQKEKKAQ-(8)-NNKKPQPSA-(9)-QYNKLAKAQTLRDKPQERQER

CteIP  CCGGG : -?-SVLP SRRKNNLQKEKKAQ-(8)-NNKKPQPSA-(9)-QYNKLAKAQTLRDKPQERQER

Cfr10I  CCGGG : -57-SDGSFNKCNKSDWEKIGRATPE-(44)-LISNDFTSOID-(40)-EHKSFEEVVRPPPGAQLHARSGSLK

BsrFI  CCGGG : -66-PASAAFSNTRTWGDKIAQSWNFEE-(47)-LISNDPILIRQ-(36)-DSAPAGKEQPPDPLNNKSHGMK

Spr2I  CCGGGG : 182-DSJAKVRDLIDDLSAGWCAA-(66)-LSJSTLAVVLI-(38)-GSLAEPETLQSTPLVWWK

NgoMIV  CCGGG : -56-GERLFGQQSTNPAECSEPKQSAFE-(51)-DYLTDITTYTRN-(31)-PLSASHCQGQAQASLGSLV

AccIII  CCGGG : 118-MQGSAKIDTVVRSHPSA-(18)-IYKCGDIIISFF-(10)-DNEIVWETKSGDRMGKFIQK

Kpn2I  CCGGG : 124-MKSGSHTPTTFELAAPVWKMI-(22)-PSVSTDVIKSRV-(1)-KNASVYLPETLERRFVFFAOQIRID

MjaVIP  CCGGG : -46-KANBEWTVTHDCQVYINELJK-LK-(35)-LLPDLADTVYKV-(4)-IKSANTVKAGFPRFETTYWKLYQ

HpIho263  CCGGG : -60-GRGFVSAVRJALGEEYEQ-(33)-ETVLQDSYIYT-(4)-IKRANATVAGFRYETETYWKLKL
Pingoud et al., Figure 4

A

R116A  R117A  R119A  K122A  E125A  D160A  K182A

bound —

free —

B


bound —

free —
Pingoud et al., Figure 5

**A**

- WT
- R116A
- R117A
- R119A
- K122A
- E125A
- D160A
- K182A

325 bp

184 bp

141 bp

**B**

R186A
E187A
R188A
E191A
E195A
E233A
K235A
WT

325 bp

184 bp

141 bp
Pingoud et al., Figure 7
Pingoud et al., Figure 8
Pingoud et al., Figure 9