Protein stability indicates divergent evolution of PD-(D/E)XK type II restriction endonucleases

MONIKA FUXREITER AND ISTVÁN SIMON

Institute of Enzymology, Hungarian Academy of Sciences, H-1518 Budapest, Pf. 7., Hungary

(RECEIVED December 19, 2001; FINAL REVISION May 8, 2002; ACCEPTED May 18, 2002)

Abstract

Type II restriction endonucleases recognize 4–8 base-pair-long DNA sequences and catalyze their cleavage with remarkable specificity. Crystal structures of the PD-(DE)XK superfamily revealed a common α/β core motif and similar active site. In contrast, these enzymes show little sequence similarity and use different strategies to interact with their substrate DNA. The intriguing question is whether this enzyme family could have evolved from a common origin. In our present work, protein structure stability elements were analyzed and compared in three parts of PD-(DE)XK type II restriction endonucleases: (1) core motif, (2) active-site residues, and (3) residues playing role in DNA recognition. High correlation was found between the active-site residues and those stabilization factors that contribute to preventing structural decay. DNA recognition sites were also observed to participate in stabilization centers. It indicates that recognition motifs and active sites in PD-(DE)XK type II restriction endonucleases should have been evolutionary more conserved than other parts of the structure. Based on this observation it is proposed that PD-(DE)XK type II restriction endonucleases have developed from a common ancestor with divergent evolution.

Keywords: Stabilization centers; DNA recognition; phosphodiester hydrolysis; structural similarity; divergent evolution

Type II restriction endonucleases catalyze phosphodiester bond hydrolysis in 4–8 base-pair-long DNA sequences (Roberts and Halford 1993). These enzymes serve as excellent model systems to study DNA recognition due to their remarkable selectivity. Type II restriction endonucleases consist of four superfamilies with distinct folds: PD-(D/E)XK, Nuc, GIY-YIG, and HNH nucleases (Aravind et al. 2000; Sapranauskas et al. 2000; Bujnicki et al. 2001). The presently solved crystal structures correspond to only one, the PD-(D/E)XK superfamily. These structures show high analogy among their active sites, which include two acidic residues and usually a Lys. In *Bam*HI and *Bgl*II, however, Lys is replaced by Glu. Also each structure contains a similar α/β core motif, which is built up by 5–6 β strands surrounded by several flanking α helices. Despite these similarities, PD-(D/E)XK endonucleases lack substantial sequence similarity. No uniform catalytic mechanism could have been established yet for type II restriction endonucleases, mostly due to the ambiguity of the metal ions involved in catalysis (Pingoud and Jeltsch 1997, 2001; Horton et al. 1998b; Viadiu and Aggarwal 1998; Horton and Cheng 2000; Fuxreiter and Osman 2001). There are several strategies described for DNA recognition by these enzymes (McClarin et al. 1986; Winkler et al. 1993; Newman et al. 1995, 1998; Horton and Cheng 2000; Lukacs et al. 2000). For example *Bam*HI and *Eco*RI, approach the DNA from the major groove, whereas *Eco*RV and *Pvu*II access DNA from the minor groove. The intriguing question is how these enzymes have evolved, whether they have a common ancestor from which they have developed. Exploring the evolutionary relationship between type II restriction endonucleases can also reveal basic principles of DNA recognition.

In this work we analyzed those factors that contribute to the structural stability of the PD-(D/E)XK superfamily of type II restriction endonucleases. These residues form co-

Reprint requests to: Monika Fuxreiter, Institute of Enzymology, Hungarian Academy of Sciences, H-1518 Budapest, Pf. 7., Hungary; e-mail: monika@enzim.hu; fax: (36-1)-466-5465.

Abbreviations: SC, stabilization center; PDB, Protein Data Bank

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.4980102.

operative sets of long-range interactions, which prevent unfolding of the structure. They are identified as stabilization centers (SCs), which have been defined and described previously (Dosztanyi et al. 1997; see also Materials and Methods).

Because mutation of SC-forming residues affects stability of the intact structure, these residues have been observed to be evolutionarily more conserved than average residues. Therefore, analyzing SCs in restriction endonucleases can identify those parts of the structure that are evolutionarily most conserved. In this study we focused on three structural–functional parts of PD-(D/E)XK endonucleases: (1) core motif, (2) active-site residues, and (3) residues playing a role in DNA recognition. Exploring common stability motifs can help to understand the evolutionary relationship between PD-(D/E)XK type II restriction endonucleases.

Results

The distribution of stabilization centers in the α/β core of *Bam*HI, *Eco*RI, *Fok*I, *Bgl*II, *Eco*RV, *Pvu*II, *Cfr*01, and *Bgl*I compared to the whole enzyme is displayed in Table 1. The core is structurally most conserved, and hence it is expected to be most stable in PD-(D/E)XK endonucleases. This motif, however, does not contain the majority of stabilization centers in each enzyme. Ratio of SCs belonging to the core covers a wide range: SCs are dominant in the core of *Eco*RV (~60%); this ratio varies between 30% and 40% in *Bgl*I and *Eco*RI, whereas in other enzymes it decreases to 10%–30%. Interestingly, *Pvu*II, which is believed to be most closely related to *Eco*RV has the lowest ratio of core SCs. In most cases the ratio of SCs in the core motif is comparable with the relative size of the core. This observation indicates that contribution of the core to structural stability of the whole enzyme is approximately proportional to the size of the core. *Eco*RV and *Pvu*II are exceptions in which the ratio of core SCs is significantly higher or lower than the ratio of the core, respectively. In *Pvu*II only 5 SCs are sufficient to keep the core intact.

The core motifs also form only a few SCs with the rest of the structure. Other structural parts are involved in more SCs, thus providing a greater contribution to preventing proteins from degradation. These results suggest that the core is important but not dominant for the structural stability of PD-(D/E)XK type II restriction endonucleases. Consequently, despite the structural similarity of the core motifs, it is unlikely that the core represents the evolutionary relationship between these enzymes.

Because the binding of DNA affects protein conformation in many enzymes, SCs in the free enzymes have been compared to complex structures. Protein–DNA complexes in general are more compact than free enzyme structures, which can increase the number of SCs. Interestingly, in *Eco*RV, *Eco*RI, and *Bgl*II, in which DNA undergoes major distortion upon interacting with the protein, the number of SCs increases in the catalytically competent complexes compared to the free enzymes. In *Bam*HI and *Pvu*II, in which DNA retains B-DNA conformation, the number of SCs in the protein remains fairly stable.

The catalytic machinery of restriction endonucleases requires the presence of several negatively charged side chains in the active site. The primary role of these residues is to ligate the catalytically essential metal ion cofactor. Correlation between SCs and active-site residues are displayed in Table 2. In all studied enzymes, at least one active-site residue is involved in an SC. The only exception is the complex of *Eco*RI with DNA, in which no active-site residue forms SC. In the free enzyme and in a catalytically active complex with a metal ion, however, two active-site residues contribute to stabilization of the structure. Involve-

Table 1. *Stabilization centers, in the core motif of type II restriction endonucleases*

			Ratio of the core (%)		Total No. of SCs			No. of SCs in the core	
Enzyme	Type	PDB	A	B	A	B	AB	A	B
*Bam*HI	free	1bam	24.9		51			14	
	DNA	1bhm	25.2	24	51	57		13	15
	DNA + Ca^{2+}	2bam	24.1	23.9	54	60	1	15	15
	DNA + Mn^{2+}	3bam	24.3	23.9	52	53	1	15	13
	DNA*	1esg	24.4	23.7	48	51	2	13	11
*Eco*RI	free	1qc9	21.6	21.6	31	31		11	11
	DNA	1eri	20.2		29			10	
	DNA + Mn^{2+}	1qps	21.1		40			12	
*Fok*I	free	2fok	9.5	9.5	85	85		15	17
	DNA	1fok	9.4		78			13	
*Bgl*II	free	1es8	29.3		49			12	
	DNA + Mg^{2+}	1d2i	26.0	26.6	55	51	3	12	11
	DNA + Ca^{2+}	1dfm	26.0	26.6	58	51	3	11	12
*Cfr*01	free	1cfr	20.1		60			11	
*Eco*RV	free	1rve	29.5	29.5	43	45	10	25	28
	DNA	4rve	30.0	30.0	44	51	11	25	27
	DNA	1eoo	32.5	32.5	56	53	15	28	25
	DNA	1eop	30.4	29.9	46	47	9	20	30
	DNA	1bgb	34.4	33.3	52	48	11	27	25
	DNA	1rva	29.5	29.5	56	55	15	29	30
	DNA + Mg^{2+}	1rvb	29.5	29.5	54	53	17	27	31
	DNA + Mg^{2+}	1rve	29.5	29.5	53	53	19	29	29
	DNA + Ca^{2+}	1az0	34.4	33.6	54	50	14	26	27
	DNA + Ca^{2+}	1b94	29.5	29.5	60	49	13	29	27
	DNA*	2rve	34.4	34.6	57	48	10	30	25
*Pvu*II	free	1pvu	38.9	38.9	31	25	12	5	5
	DNA	1eyu	38.5	39.2	28	32	10	7	7
	DNA	1pvi	38.9	38.9	19	24	3	3	4
	DNA	3pvi	38.9	38.9	25	23	10	3	3
	DNA + Ca^{2+}	1f0o	38.9	39.5	26	32	2	5	8
*Bgl*I	DNA + Ca^{2+}	1dmu	28.3		61			25	

A and B stand for the subunits. AB designates stabilization centers between two subunits. If the enzyme is in monomer form, all SCs are listed for the A subunit. DNA* means complex with nonspecific substrate.

Table 2. *Active site residues participating in stabilization centers of type II restriction endonucleases*

Enzyme	Active-site residues	Type	PDB	Ref	SCs with active-site residues
*Bam*HI	Glu 77, Asp 94, Glu 111, Glu 113	free	1bam	1	**Glu 111**–Ile 141
		DNA	1bhm	2	**Asp 94**–Met 110, **Glu 111**–Ile 141
		DNA + Ca^{2+}	2bam	3	**Asp 94**–Met 110, **Glu 111**–Ile 141
		DNA + Mn^{2+}	3bam	3	**Asp 94**–Met 110, **Glu 111**–Ile 141
		DNA*	1esg	4	**Asp 94**–Met 110, **Glu 111**–Ile 141
*Eco*RI	Glu 37, Asp 91, Glu 111, Lys 113	free	1qc9	5	**Asp 91**–Ala 110, **Asp 91–Glu 111, Glu 111**–Leu 167
		DNA	1eri	6	
		DNA + Mn^{2+}	1qps	7	**Asp 91**–Ala 110, **Asp 91–Glu 111, Glu 111**–Leu 167
*Bgl*II	Asn 69, Asp 84, Glu 93, Gln 95	free	1es8	8	**Asn 96–Asp 84, Asn 69**–Ile 83, **Glu 93**–Ile 129, **Gln 95**–Ile 130
		DNA + Mg^{2+}	1d2i	9	**Glu 93**–Ile 129, **Gln 95**–Ile 130
		DNA + Ca^{2+}	1dfm	9	**Glu 93**–Ile 129, **Gln 95**–Ile 130
*Fok*I	Glu 425, Asp 450, Asp 467, Lyd 469	free	2fok	10	**Asp 450**–Val 466, Asp 421–**Lys 469**
		DNA	1fok	11	**Asp 450**–Val 466
*Cfr*10I	Glu 71, Asp 134, Ser 188, Lys 190	free	1cfr	12	**Asp 134**–Leu 187, **Asp 134–Ser 188, Ser 188**–Ala 231, **Lys 190**–Ala 232
*Eco*RV	Glu 45, Asp 74, Asp 90, Lys 92	Free	1rve	13	**Asp 74**–Ile 89, **Asp 74–Asp 90, Asp 90**–Ile 133
		DNA	4rve	13	**Asp 74–**Ile 89, **Asp 74–Asp 90, Asp 90**–Ile 133
		DNA	1eoo	14	**Asp 74–**Ile 89, **Asp 74–Asp 90, Asp 90**–Ile 133
		DNA	1eop	14	**Asp 74–**Ile 89, **Asp 74–Asp 90, Asp 90**–Ile 133
		DNA	1bgb	15	**Asp 74–**Ile 89, **Asp 74–Asp 90, Asp 90**–Ile 133
		DNA	1rva	13	**Asp 74–**Ile 89, **Asp 74–Asp 90, Asp 90**–Ile 133
		DNA + Mg^{2+}	1rvb	16	**Asp 74–**Ile 89, **Asp 74–Asp 90, Asp 90**–Ile 133
		DNA + Mg^{2+}	1rvc	16	**Asp 74–**Ile 89, **Asp 74–Asp 90, Asp 90**–Ile 133
		DNA + Ca^{2+}	1az0	17	**Asp 74–**Ile 89, **Asp 74–Asp 90, Asp 90**–Ile 133
		DNA + Ca^{2+}	1b94	18	**Asp 74–**Ile 89, **Asp 74–Asp 90, Asp 90**–Ile 133
		DNA*	2rve	13	**Asp 74–Asp 90**
*Pvu*II	Glu 55, Asp 58, Glu 68, Lys 70	free	1pvu	19	**Glu 68**–Val 97, **Glu 68**–Pro 98, **Glu 68**–Trp 99
		DNA	leyu	20	**Glu 68**–Pro 98, **Glu 68**–Trp 99
		DNA	lpvi	21	**Glu 68**–Pro 98, **Glu 68**–Trp 99
		DNA	3pvi	22	**Glu 68**–Pro 97, **Glu 68**–Pro 98, **Glu 68**–Trp 99
		DNA + Ca^{2+}	1f0o	20	**Glu 68**–Pro 97, **Glu 68**–Pro 98, **Glu 68**–Trp 99
*Bgl*I	Glu 87, Asp 116, Asp 142, Lys 144	DNA + Ca^{2+}	1dmu	23	**Asp 116**–Val 141
*Mun*I	Asp 83, Glu 98, Lys 100, Arg 21	DNA	1d02	24	**Ala 83**–Gly 97, **Ala 83–Glu 98,** Phe 84–**Glu 98**
*Nae*I	Glu 70, Asp 86, Asp 95, Lys 97	free	1ev7	25	**Asp 95**–Cys 116, **Asp 95**–Leu 117, **Lys 97**–Leu 117
		DNA	1iaw	26	**Asp 95**–Ile 115, **Asp 95**–Cys 116, **Asp 95**–Leu 117
*Ngo*MIV	Glu 70, Asp 140, Glu 201, Lys 187	DNA + Mg^{2+}	1fiu	27	**Asp 140**–Ile 184
*Bso*BI	Asp 212, Glu 240, Lys 242, His 253	DNA	1dc1	28	**Asp 212**–Gly 238, **Asp 212–Glu 240**
*Bse*634I	Asp 146, Lys 198, Glu 212, Glu 80	DNA	1knv	29	**Asp 146**–Val 195
*Hinc*II	Asp 114, Asp 127, Asp 129**	DNA + Na^{+}	1kc6	30	**Asp 114**–Leu 126, **Asp 127**–Tyr 168, **Asp 127**–Leu 169, **Lys 129**–Glu 170

The active-site residues are marked in bold. DNA* means complex with nonspecific substrate. **The active-site residues of *Hin*dII are based on superposition with the *Eco*RV structure (N. Horton, pers. comm.)

1. Newman et al. 1994; 2. Newman et al. 1995; 3. Viadiu and Aggarwal 1998; 4. Viadiu and Aggarwal 2000; 5. Kim et al. 1990; 6. McClarin et al. 1986; 7. Horvath et al. 1999; 8. Lukacs et al. 2001; 9. Lukacs et al. 2000; 10. Wah et al. 1998; 11. Wah et al. 1997; 12. Bozic et al. 1996; 13. Winkler et al. 1993; 14. Horton and Perona 2000; 15, Horton and Perona 1998; 16. Kostrewa and Winkler 1995; 17. Perona and Martin 1997; 18. Thomas et al. 1999; 19. Athanasiadis et al. 1994; 20. Horton and Cheng 2000; 21. Cheng et al. 1994; 22. Horton et al. 1998; 23. Newman et al. 1998; 24. Deibert et al. 1999; 25. Huai et al. 2000; 26. Huai et al. 2001; 27. Deibert et al. 2000; 28. van der Woerd et al. 2001; 29. Grazulis et al. 2002; 30. Horton et al. 2002.

ment of two or more active-site residues in SC elements is quite frequent. Interestingly, in several cases, like *Eco*RI, *Eco*RV, *Cfr*10I, *Bgl*II, *Mun*I, and *Bso*BI two active-site residues form an SC with each other. In *Bam*HI complexes and in *Fok*I a residue next to an active-site residue makes SC link with another active-site residue. It suggests that the active-site residues provide an important contribution to the stabilization of the whole enzyme structure. In other words, although the active-site residues belong to different secondary structure elements, they form a structurally stable unit, which should have been conserved during evolution. The fact that not all active-site residues are involved in SC formation can explain the small variability of the active sites in PD-(D/E)XK type II restriction endonucleases.

Type II restriction endonucleases use diverse strategies to interact with their substrate DNA (Aggarwal 1995). Several attempts have been made to establish a correlation between the specific base-pair sequences and the recognition motifs. Besides understanding this fascinating phenomena, there is a practical goal behind these works, namely to alter the

specificity of restriction endonucleases. During the SC analysis of restriction endonucleases, a correlation has been found between some of the recognition residues and the stabilization elements. Those residues, which play a role in DNA recognition and are also involved in SCs, are shown in Table 3. For all studied type II endonucleases, with the exception of *Ngo*MIV, at least one residue binding to the cognate DNA sequence also plays a role in SC formation. In the case of *Ngo*MIV Ser 40, which orients a recognition residue, Arg 227, participates in SC formation. The presence of SCs indicates an extensive set of interactions between two parts of the structure, which are responsible for making specific contacts with DNA. It is also reflected by the fact that in several cases the neighboring side chains of a recognition residue are involved in SC formation. At first, the contribution of recognition residues to protein stability is surprising. Because restriction endonucleases work on more than 200 sequences (Roberts and Macelis 2001), the recognition motifs are expected to be highly variable during evolution. On the other hand, our observation means that bringing together those residues, which are responsible for specific DNA binding, also contributes to stabilization of the structure. Thus, it makes the recognition motif evolutionarily more conserved. It means, that PD-(D/E)XK endonucleases contain a stable motif with appropriate geometry to contact with either or both DNA grooves. This structural element is held together with an extensive set of long-range interactions around the central residue, which is

Table 3. *Recognition sites participating in stabilization centers of type II restriction endonucleases*

Enzyme	Type	PDB	SCs with recognition sites
*Bam*HI	free	1bam	**Asn 116–Val 156,** Thr 144–**Val 156,** *Gly 115*–**Val 156,** Asn 55–**Thr 153,** Gly 56–**Thr 153**
	DNA	1bhm	**Asn 116–Val 156,** *Gly 115*–**Val 156,** *Ile 117*–**Val 156, Asn 116**–*Thr 157,* Asn 53–**Thr 153,** Asn 55–**Thr 153,** Gly 56–**Thr 153**
	DNA + Ca^{2+}	2bam	**Asn 116–Val 156, Asn 116**–Thr 157, *Ile 117*–**Val 156,** *Gly 115*–**Val 156,** Asn 55–**Thr 153,** Gly 56–**Thr 153**
	DNA + Mn^{2+}	3bam	**Asn 116–Val 156,** *Ile 117*–**Val 156, Asn 116**–*Thr 157,* Asn 53–**Thr 153,** Asn 55–**Thr 153,** Gly 56–**Thr 153**
	DNA*	1esg	Gly 115–**Val 156**
*Eco*RI	free	1qc9	**Asn 141–Arg 203**
	DNA	1eri	*Ile 143*–**Arg 203,** *Glu 144*–**Arg 203,** Phe 174–**Arg 200**
	DNA + Mn^{2+}	1qps	*Ile 143*–**Arg 203,** Phe 174–**Arg 200**
*Fok*I	free	2fok	**Gln 12–Arg 228,** Phe 178–**Glu 220**
	DNA	1fok	Phe 178–**Glu 220**
*Bgl*II	free	1es8	**Ser 97**–*Ser 142*
	DNA + Mg^{2+}	1d2i	**Asn 98**–*Ser 142,* **Asn 98**–Leu 143
	DNA + Ca^{2+}	1dfm	**Asn 98**– Leu 143
*Eco*RV	free	1rve	**Thr 106**–Gly 190, Gly 108–**Asn 188,** Gly 109–**Asn 188,** Tyr 110–**Asn 188**
	DNA	4rve	**Thr 106**–*Ile 189,* **Thr 106**–Gly 190, *Leu 107*–**Asn 188,** Gly 108–**Asn 188,** Gly 109–**Asn 188,** Tyr 110–**Asn 188**
	DNA	1eoo	**Thr 106**–*Ile 189,* **Thr 106**–Gly 190, *Leu 107*–**Asn 188,** Gly 108–**Asn 188,** Gly 109–**Asn 188,** Tyr 110–**Asn 188**
	DNA	1eop	**Thr 106**–*Ile 189,* **Thr 106**–Gly 190, *Leu 107*–**Asn 188,** Gly 108–**Asn 188,** Gly 109–**Asn 188,** Tyr 110–**Asn 188**
	DNA	1bgb	**Thr 106**–*Ile 189,* **Thr 106**–Gly 190, *Leu 107*–**Asn 188,** Gly 108–**Asn 188,** Gly 109–**Asn 188,** Tyr 110–**Asn 188**
	DNA	1rva	**Thr 106**–*Ile 189,* **Thr 106**–Gly 190, *Leu 107*–**Asn 188,** Gly 108–**Asn 188,** Gly 109–**Asn 188,** Tyr 110–**Asn 188**
	DNA + Mg^{2+}	1rvb	**Thr 106**–*Ile 189,* **Thr 106**–Gly 190, Gly 108–**Asn 188,** Gly 109–**Asn 188,** Tyr 110–**Asn 188**
	DNA + Mg^{2+}	1rvc	**Thr 106**–*Ile 189, Leu 107*–**Asn 188,** Gly 108–**Asn 188,** Gly 109–**Asn 188,** Tyr 110–**Asn 188**
	DNA + Ca^{2+}	1az0	**Thr 106**–*Ile 189,* **Thr 106**–Gly 190, *Leu 107*–**Asn 188,** Gly 108–**Asn 188,** Gly 109–**Asn 188,** Tyr 110–**Asn 188**
	DNA + Ca^{2+}	1b94	**Thr 106**–*Ile 189,* **Thr 106**–Gly 190, *Leu 107*–**Asn 188,** Gly 108–**Asn 188,** Gly 109–**Asn 188,** Tyr 110–**Asn 188**
	DNA*	2rve	**Thr 106**–Gly 190, *Leu 107*– **Asn 188,** Gly 108–**Asn 188,** Gly 109–**Asn 188,** Tyr 110–**Asn 188**
*Pvu*II	free	1pvu	**His 84–Asn 141, Thr 82–Asn 141, His 83–Asn 141, Thr 82**–*Pro 142,* **His 84–Asn 141**
	DNA	1eyu	**Ser 81**–*Lys 143,* **Thr 82**–*Pro 142,* **His 83–Asn 141, His 84–Asn 141**
	DNA	1pvi	**Ser 81**–*Lys 143,* **Thr 82**–*Pro 142,* **His 83–Asn 141**
	DNA	3pvi	**Ser 81**–*Lys 143,* **Thr 82**–*Pro 142,* **His 83–Asn 141, His 84–Asn 141**
	DNA + Ca^{2+}	1f0o	**His 83–Asn 141, His 83–Asn 140, His 84–Asn 140, His 84–Asn 141**
*Bgl*I	DNA + Ca^{2+}	1dmu	**Asp 154**–*Val 280,* **Asp 154**–*Asp 281, Leu 155*–**Arg 279,** Val 156–*Val 278, Val 156*–**Arg 279,** Arg 263–**Arg 279,** Pro 264–**Arg 279**
*Mun*I	DNA	1d02	Lys 50–**Gly 79,** Asn 51–**Gly 79,** Leu 52–**Gly 79,** Tyr 53–**Gly79, Ile 99**–Val 151, **Asp 103**–Asp 155, **Asp 103**–Ile 156
*Nae*I	free	1ev7	**Phe 98**–Trp 120, **Ser 99**–Trp 120
	DNA	1iaw	**Phe 98**–Trp 120, **Ser 99**–Trp 120
*Ngo*MIV	DNA + Mg^{2+}	1fiu	Ser 40–Thr 224
*Bso*BI	DNA	1dcl	Tyr 24–**Asp 246,** Phe 28–**Asp 246**
*Bse*634I	DNA	1knv	Ile 149–**Ala 193, Ala 193**–Lys 235, **Gly 194**–Lys 235, **Gly 194**–Tyr 236
*Hinc*II	DNA + Na^{+}	1kc6	**Asn 141**–Ile 208, **Asn 141**–Gln 209, **Ala 139**–Phe 210, Pro 140–**Gln 209,** Asn 141–**Gln 209**

Residues that interact with DNA are marked in bold. Neighboring residues are marked in italic. DNA* means complex with nonspecific substrate.

important for the stability of the protein structure. The structural stability has been proposed to contribute to preserving this motif and hence its DNA recognition function during evolution. As there is only one conserved pair in the interaction of the recognition residues, the rest is highly variable, providing diversity in the cognate sequences.

Type II restriction endonucleases can be divided into α and β subfamilies (Huai et al. 2000; Bujnicki 2001), which have different dimerization schemes and DNA recognition patterns. Enzymes belonging to the α family (e.g., *Eco*RI, *Bam*HI, *Mun*I, *Fok*I, and *Cfr*10I) use an α helix and a loop for DNA recognition and usually produce 5′ overhanging ends, whereas members of the β family (e.g., *Eco*RV, *Pvu*II, *Bgl*I, and *Nae*I) use a β strand to interact with the cognate DNA sequence and give blunt-ended products. It is interesting to analyze whether SCs indicate any difference between these two endonuclease subfamilies. Residues involved in dimerization were observed to form very few SC links; hence they cannot be used as a distinction criteria. Those SC-forming residues that participate in DNA recognition, however, interact with different base pairs in the two subfamilies. In enzymes of the α family almost all SC-forming recognition residues make contact with either or both of the outer two guanines, whereas in the β family, they interact with the second thymine. The present definition of SCs, however, is not suitable for quantitative characterization of evolutionary distances of enzymes. Hence, based on SC similarities, we cannot derive a structure-based phylogenetic tree of restriction endonucleases (Bujnicki 2000).

The SC analysis was also extended to two related DNA repair nucleases: the mismatch repair *Mut*H (Ban and Yang 1998) and the very short patch repair *Vsr* endonuclease (Tsutakawa et al. 1999), which exhibit the same fold as PD-(D/E)XK endonucleases. In *Mut*H two active-site residues (Glu 77 and Lys 79) and one recognition residue (Phe 94) participate in SC formation. In *Vsr* Asp 51 of the active site and Gly 65, as well as Glu 116 of the recognition motif, are involved in SCs. These results indicate that some of the active-site and recognition residues contribute to overall structural stability in all nucleases with PD-(D/E)XK fold.

Discussion

Stabilization centers have been analyzed to test the possible evolutionary relationships between structures of PD-(D/E)XK type II endonucleases. In general, the common α/β core motif was not found to provide a major contribution to structure stabilization in these enzymes. The active-site residues, as well as some residues in the recognition sites, are, however, persistently involved in SC formation. Hence, these parts of the enzyme—the active site and the recognition site–can be concluded to be evolutionarily most conserved in PD-(D/E)XK endonucleases. Conserving a stabilization center in these sites, however, does not prohibit the variability of these sites. In most cases two residues of the active site or recognition site are involved in SC formation. The conservation of other residues is not required to provide sufficient stability for these structural motifs via an extensive set of long range interactions. It can explain the diversity of the DNA sequences, which can be recognized by these enzymes. Our results support the hypothesis that PD-(D/E)XK type II restriction endonucleases have been developed from a common ancestor with divergent evolution.

Materials and methods

Crystal structures of 14 enzymes: *Bam*HI, *Eco*RI, *Eco*RV, *Pvu*II, *Bgl*I, *Bgl*II, *Fok*I, *Cfr*10I, *Mun*I, *Nae*I, *Ngo*MIV, *Bso*BI, *Bse*634I, and *Hinc*II were analyzed in free forms, in complexes with substrate DNA, and in catalytically active form with DNA and metal ions. The Protein Data Bank (PDB) codes are displayed in Tables 1–3; the references are given in Table 2. *Bam*HI and *Eco*RV complexes with nonspecific DNA were also included. The crystal structures were not optimized. The core motifs in Table 1 were defined to include five β strands and two α helices involved in dimerization, which are as follows: *Bam*HI (β^3, β^4, β^5, β^6, β^7, α^4, and α^6); *Eco*RI (β^1, β^2, β^3, β^4, β^5, α^4, and α^5); *Eco*RV (β^c, β^d, β^e, β^g, β^h, α^A, and α^B); *Pvu*II (β^a, β^b, β^c, β^e, β^f, α^A, and α^B); *Bgl*I (β^1, β^2, β^3, β^8, β^9, α^2, and α^4); *Bgl*II (β^3, β^4, β^5, β^6, β^7, α^4, and α^5); *Fok*I (β^1, β^2, β^3, β^4, β^5, α^4, and α^5); and *Cfr*10I (β^3, β^4, β^5, β^6, β^7, α^7, and α^8).

The stabilization centers were calculated using the original definition (Dosztanyi et al. 1997). Two residues form an SC element if (1) they are involved in long-range interaction, that is, they are separated by at least 10 residues in sequence and the contact distance of their two closest atoms is less than the sum of their van der Waals radii plus 1 Å, and (2) two supporting residues can be selected from both of their flanking tetrapeptides, which together with the central residues form at least seven out of the possible nine contacts.

Acknowledgments

We thank Prof. Roman Osman for stimulating discussions. The research has been sponsored by OTKA grants T30566, T34131, U.S.-Hungarian Mobility grant 99/MO/01, as well as Bolyai and OTKA D34572 fellowships.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Aggarwal, A.K. 1995. Structure and function of restriction endonucleases. *Curr. Opin. Struct. Biol.* **5:** 11–19.

Aravind, L., Makarova, K., and Koonin, E.V. 2000. Holliday junction resolvases and related nucleases: Identification of new families, phyletic distribution, and evolutionary trajectories. *Nucleic Acids Res.* **28:** 3417–3432.

Athanasiadis, A., Vlassi, M., Kotsifaki, D., Tucker, P.A., Wilson, K.S., and Kokkinidis, M. 1994. Crystal structure of PvuII endonuclease reveals extensive structural homologies to EcoRV. *Nat. Struct. Biol.* **1:** 469–475.

Ban, C. and Yang, W. 1998. Structural basis for *Mut*H activation in *E. coli* mismatch repair and relationship of *Mut*H to restriction endonucleases. *EMBO J.* **17:** 1526–1534.

Bozic, D., Grazulis, S., Siksnys, V., and Huber, R. 1996. Crystal structure of Citrobacter freundii restriction endonuclease Cfr10I at 2.15 Å resolution. *J. Mol. Biol.* **255:** 176–186.

Bujnicki, J. 2000. Phylogeny of the restriction endonuclease superfamily inferred from comparison of protein structures. *J. Mol. Evol.* **50:** 39–44.

———. 2001. Understanding the evolution of restriction-modification systems: Clues from sequence and structure comparisons. *Acta Biochim. Pol.* **48:** 935–967.

Bujnicki, J., Radlinska, M., and Rychlewski, L. 2001. Polyphyletic evolution of type II restriction enzymes revisited: Two independent sources of second-hand folds revealed. *Trends Biochem. Sci.* **26:** 9–11.

Cheng, X., Balendiran, K., Schildkraut, I., and Anderson, J.E. 1994. Structure of PvuII endonuclease with cognate DNA. *EMBO J.* **13:** 3927–3935.

Deibert, M., Grazulis, S., Arvydas, J., Siksnyis, V., and Huber, R. 1999. Crystal structure of MunI restriction endonuclease in complex with cognate DNA at 1.7 Å resolution. *EMBO J.* **18:** 5805–5816.

Deibert, G., Grazulis, S., Sasnauskas, G., Siksnys, V., and Huber, R. 2000. Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. *Nat. Struct. Biol.* **7:** 792–799.

Dosztanyi, Z., Fiser, A., and Simon, I. 1997. Stabilization centers in proteins: Identification, characterization, and predictions. *J. Mol. Biol.* **272:** 597–612.

Fuxreiter, M., and Osman, R. 2001. Probing the general base catalysis in the first step of *Bam*HI action by computer simulations. *Biochemistry* **40:** 15017–15023.

Grazulis, S., Deibert, M., Rimseliene, R., Skirgaila, R., Sasnauskas, G., Lagunavicius, A., Repin, V., Urbanke, C., Huber, R., and Siksnys, V. 2002. Crystal structure of the Bse634I restriction endonuclease: Comparison of two enzymes recognizing the same DNA sequence. *Nucleic Acids Res.* **30:** 876–885.

Horton, J.R. and Cheng, X. 2000. *Pvu*II endonuclease contains two calcium ions in active sites. *J. Mol. Biol.* **300:** 1049–1056.

Horton, J.R., Nastri, H.G., Riggs, P.D., and Cheng, X. 1998. Asp34 of PvuII endonuclease is directly involved in DNA minor groove recognition and indirectly in catalysis. *J. Mol. Biol.* **284:** 1491–1504.

Horton, N., Dorner, L., and Perona, J. 2002. Sequence selectivity and degeneracy of a restriction endonuclease mediated by DNA intercalation. *Nat. Struct. Biol.* **9:** 42–47.

Horton, N.C. and Perona, J.J. 1998. Recognition of flanking DNA sequences by EcoRV endonuclease involves alternative patterns of water-mediated contacts. *J. Biol. Chem.* **273:** 21721–21729.

Horton, N.C., Newberry, K.J., and Perona, J.J. 1998. Metal ion-mediated substrate-assisted catalysis in type II restriction endonucleases. *Proc. Natl. Acad. Sci.* **95:** 13489–13494.

Horton, N.C., and Perona, J.J. 2000. Crystallographic snapshots along protein induced DNA bending pathway. *Proc. Natl. Acad. Sci.* **97:** 5729–5734.

Horvath, M.M., Choi, J., Kim, Y., Wilkosz, P., and Rosenberg, J.M. 1999. The integration of recognition and cleavage: X-ray structures of pretransition state and post reactive DNA-Eco RI endonuclease complexes (in press).

Huai, Q., Colandene, J.D., Chen, Y., Luo, F., Zhao, Y., Topal, M.D., and Ke, H. 2000. Crystal structure of *Nae*I—An evolutionary bridge between DNA endonuclease and topoisomerase. *EMBO J.* **19:** 3110–3118.

Huai, Q., Colandene, J., Topal, M., and Ke, H. 2001. Structure of NaeI-DNA complex reveals dual-mode DNA recognition and complete dimer rearrangement. *Nat. Struct. Biol.* **8:** 665–669.

Kim, Y.C., Grable, J.C., Love, R., Greene, P.J., and Rosenberg, J.M. 1990. Refinement of Eco RI endonuclease crystal structure: A revised protein chain tracing. *Science* **249:** 1307–1309.

Kostrewa, D. and Winkler, F.K. 1995. Mg^{2+} binding to the active site of EcoRV endonuclease: A crystallographic study of complexes with substrate and product DNA at 2 Å resolution. *Biochemistry* **34:** 683–696.

Lukacs, C.M., Kucera, R., Schildkraut, I., and Aggarwal, A.K. 2000. Understanding the immutability of restriction enzymes: Crystal structure of *Bgl*II and its DNA substrate at 1.5 Å resolution. *Nat. Struct. Biol.* **7:** 134–140.

———. 2001. Structure of free BglII reveals an unprecedented scissor-like motion for opening an endonuclease. *Nat. Struct. Biol.* **8:** 126–130.

McClarin, J.A., Frederick, C.A., Wang, B.C., Greene, P., Boyer, H.W., Grable, J., and Rosenberg, J.M. 1986. Structure of the DNA-*Eco*RI endonuclease recognition complex at 3 Å resolution. *Science* **234:** 1526–1541.

Newman, M., Lunnen, K., Wilson, G., Greci, J., Schildkraut, I., and Phillips, S.E. 1998. Crystal structure of restriction endonuclease *Bgl*I bound to its interrupted DNA recognition sequence. *EMBO J.* **17:** 5466–5476.

Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I., and Aggarwal, A.K. 1994. Structure of restriction endonuclease BamIII and its relationship to EcoRI. *Nature* **368:** 660–664.

———. 1995. Structure of *Bam*HI endonuclease bound to DNA: Partial folding and unfolding on DNA binding. *Science* **269:** 656–663.

Perona, J.J. and Martin, A.M. 1997. Conformational transitions and structural deformability of EcoRV endonuclease revealed by crystallographic analysis. *J. Mol. Biol.* **273:** 207–225.

Pingoud, A. and Jeltsch, A. 1997. Recognition and cleavage of DNA by type-II restriction endonucleases. *Eur. J. Biochem.* **246:** 1–22.

———. 2001. Structure and function of type II restriction endonucleases. *Nucleic Acids Res.* **29:** 3705–3727.

Roberts, R. and Macelis, D. 2001. REBASE-restriction enzymes and methylases. *Nucleic Acids Res.* **29:** 268–269.

Roberts, R.J. and Halford, S.E. 1993. Type II restriction endonucleases. In *Nucleases* (eds. S.M. Linn et al.), pp. 35–88. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sapranauskas, R., Sasnauskas, G., lagunavicius, A., Vilkaitis, G., Lubys, A., and Siksnys, V. 2000. Novel subtype of type IIs restriction enzymes. *Bfi*I endonuclease exhibits similarities to the EDTA-resistant nuclease *Nuc* of *Salmonella typhimuruim*. *J. Biol. Chem.* **275:** 30878–30885.

Thomas, M.P., Brady, R.L., Halford, S.E., Sessions, R.B., and Baldwin, G.S. 1999. Structural analysis of a mutational hot-spot in the EcoRV restriction endonuclease: A catalytic role for a main chain carbonyl group. *Nucleic Acids Res.* **27:** 2438–3445.

Tsutakawa, S.E., Muto, T., Kawate, T., Jingami, H., Kunishima, N., Ariyoshi, M., Kohda, D., Nakagawa, M., and Morikawa, K. 1999. Crystallographic and functional studies of very short patch repair endonuclease. *Mol. Cell* **3:** 621–628.

van der Woerd, M., Pelletier, J., Xu, S., and Friedman, A. 2001. Restriction Enzyme BsoNI-DNA Complex: A tunnel for recognition of degenerate DNA sequences and potential histidine catalysis. *Structure* **9:** 133–144.

Viadiu, H. and Aggarwal, A.K. 1998. The role of metals in catalysis by the restriction endonuclease *Bam*HI. *Nat. Struct. Biol.* **5:** 910–916.

———. 2000. Structure of BamHI bound to nonspecific DNA: A model for DNA sliding. *Mol. Cell* **5:** 889–895.

Wah, D.A., Hirsch, J.A., Dorner, L.F., Schildkraut, I., and Aggarwal, A.K. 1997. Structure of the multimodular endonuclease FokI bound to DNA. *Nature* **388:** 97–100.

Wah, D.A., Bitinaite, J., Schildkraut, I., and Aggarwal, A.K. 1998. Structure of FokI has implications for DNA cleavage. *Proc. Natl. Acad. Sci.* **95:** 10564–10569.

Winkler, F.K., Banner, D.W., Oefner, C., Tsernoglou, D., Brown, R.S., Heathman, S.P., Bryan, R.K., Martin, P.D., Petratos, K., and Wilson, K.S. 1993. The crystal structure of *Eco*RV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.* **12:** 1781–1795.