J Mol Evol (2000) 50:39–44
DOI: 10.1007/s002399910005

JOURNAL OF MOLECULAR EVOLUTION

Phylogeny of the Restriction Endonuclease-Like Superfamily Inferred from Comparison of Protein Structures

Janusz M. Bujnicki

Molecular Biology Research Program. Henry Ford Health System, One Ford Place, Suite 5D, Detroit, MI 48202, USA

Received: 14 June 1999 / Accepted: 11 August 1999

Abstract. To date all attempts to derive a phyletic relationship among restriction endonucleases (ENases) from multiple sequence alignments have been limited by extreme divergence of these enzymes. Based on the approach of Johnson et al. (1990), I report for the first time the evolutionary tree of the ENase-like protein superfamily inferred from quantitative comparison of atomic coordinates of structurally characterized enzymes. The results presented are in harmony with previous comparisons obtained by crystallographic analyses. It is shown that λ-exonuclease initially diverged from the common ancestor and then two "endonucleolytic" families branched out, separating "blunt end cutters" from "5′ four-base overhang cutters." These data may contribute to a better understanding of ENases and encourage the use of structure-based methods for inference of phylogenetic relationship among extremely divergent proteins. In addition, the comparison of three-dimensional structures of ENase-like domains provides a platform for further clustering analyses of sequence similarities among different branches of this large protein family, rational choice of homology modeling templates, and targets for protein engineering.

Key words: Restriction endonucleases — Divergent evolution — Distance matrix methods — Phyletic trees from X-ray crystal structures — Target recognition — Remote homology

Correspondence to: Janusz M. Bujnicki; *e-mail:* iamb@ibbrain.ibb.waw.pl or jbujnicl@hfhs.org

Introduction

Restriction endonucleases (ENases) are a large and diverse family of enzymes which, in the presence of Mg^{2+}, recognize short DNA sequences (typically 4–8 bp long) and cleave the target in both strands, leaving 5′-phosphate and 3′-hydroxyl groups (reviewed by Bickle and Kruger 1993). Based on cofactor requirements, site specificity, and enzymatic mechanism, they have been classified into five types: I, II (with a subclass IIs), III, IV, and *Bcg*I-like (Pingoud and Jeltsch 1997). ENases are considered to be a part of "xenophobic" restriction-modification (RM) systems, ubiquitous among Bacteria and Archaea and serving to protect the cell against invading DNA (Jeltsch and Pingoud 1996).

Type-II ENases are intensely studied enzymes from a structure–function perspective. Because of their high specificity of cleavage at DNA recognition sites, they not only are among the most often used enzymes in genetic engineering, but also serve as model systems for analyzing protein–DNA interactions (Pingoud and Jeltsch 1997). More than 2500 type-II ENases have been isolated (Roberts and Macelis 1999) and crystal structures of the *Eco*RI (Kim et al. 1990), *Eco*RV (Winkler et al. 1993), *Bam*HI (Newman et al. 1994), *Pvu*II (Cheng et al. 1995), *Cfr*10I (Bozic et al. 1996), and *Bgl*I (Newman et al. 1998) ENases have been obtained. Comparisons of these structures revealed a common structural core (a five-stranded mixed β-sheet flanked by α-helices) and a very weakly conserved catalytic sequence motif, $PDX_{10–30}$(D/E)XK—evidence of an extremely distant common ancestor (Venclovas et al. 1994; Aggarwal 1995).

Recent structural and mutational analysis of *Cfr*10I has suggested that sequence similarities between ENases have eroded by way of divergent evolution to such a degree that spatial side-chain conservation plays the dominant role and the linear position of conserved active site residues is less important (Bozic et al. 1996; Skirgaila et al. 1998). Many ENases also exhibit variations in the three-dimensional arrangement of structural elements flanking the common core (Pingoud and Jeltsch 1997). This precludes straightforward use of ENase amino acid sequences in phylogenetic inference but also implies that ENases not possessing the "(D/E)XK" motif also might exhibit similar active-site architecture (Skirgaila et al. 1998). For many type II ENases, for which only the primary sequence is available, secondary structure predictions and threading results suggest that this is indeed the case, however, several enzymes seem to possess a different three-dimensional fold (J. Bujnicki and L. Rychlewski, manuscript in preparation).

The latest X-ray crystallography results of type IIs ENase *Fok*I (Wah et al. 1997), mismatch repair ENase MutH (Ban and Yang 1998), and bacteriophage λ exonuclease (λ-exo) (Kovall and Matthews 1998) provided strong evidence that, despite differences in function and overall fold, the catalytic domains of these enzymes share a similar core and presumably also the catalytic mechanism with type II ENases.

To date all attempts to infer a general sequence–structure–specificity relationship between ENases and their homologues have been restricted by the extreme divergence between sequences of the enzymes distinguished by their target specificity (Jeltsch et al. 1995). Guided by structural and biochemical studies, a common reaction mechanism of phosphodiester bond cleavage has been suggested (Jeltsch et al. 1992; Horton et al. 1998). Clustering analysis based on similarities of both amino acid sequences and target DNA specificities indicated that the majority of strongly diverged ENases share a common ancestor (Jeltsch et al. 1995). Nonetheless, study of molecular phylogeny of ENases has been limited, and despite some success in evolutionary analysis of sparse groups of closely related enzymes (predominantly isoschizomers: ENases with identical recognition and cleavage specificity), the general relationships within the superfamily are still unclear (Kroger et al. 1984; Withers et al. 1992; Siksnys et al. 1995; Jeltsch et al. 1995). It has even been concluded that, on the basis of sequence comparisons alone, an evolutionary relationship of the ENases cannot be derived (Jeltsch et al. 1995).

To investigate the phylogeny of ENases and to compare evolutionary relationships with the present-day distribution of functional features among these enzymes (e.g., specificity), another approach was necessary. The elegant pioneering work by Johnson et al. (1990) demonstrated that the quantitative measure of dissimilarity of the three-dimensional position of the main-chain atoms of homologous proteins generally can be applied to infer evolutionary trees, showing a linear relationship between the structure- and the sequence-based distance scores. The authors concluded that in cases where protein sequence comparisons are statistically insignificant or unreliable, the evolutionary relationship can still be inferred from more informative calculations of fractions of equivalent, superimposable residues, and root mean square (RMS; the square root of the average square Euclidean distances over all equivalent pairs of α-carbon positions) distances between α-carbons of aligned homologous structures. Recently, variants of this method have been successfully applied to infer the evolutionary history of other families of remotely related proteins, AdoMet-dependent methyltransferases (Bujnicki 1999) and periplasmic binding proteins (Fukami-Kobayashi et al. 1999).

Materials and Methods

Alignment of Structures and Sequences

All enzymes exhibiting restriction ENase-like fold analyzed in this study are listed in Table 1. The structures were obtained from the Protein Data Bank (PDB) (Bernstein et al. 1977), except for that of *Bgl*I, kindly provided by Dr. Simon Phillips. The atomic coordinates of proteins were superimposed using an automated protein structure alignment server at http://cape6.scripps.edu/baohong/3Dcomp/ (Zhang and Ding 1993) and a fixed cutoff rate to define equivalent α-carbon positions. For the final calculations a 3.5-Å cutoff rate was used, however, all structure-based alignments remained virtually identical for a range of values as wide as 3.0–3.8 Å, while only the fraction of α-carbon pairs below the cutoff varied (data not shown). All pairwise superimpositions showed absolute agreement with results obtained from the DALI server, using a different structure superposition algorithm with a floating RMS cutoff rate (Holm and Sander 1997), and therefore less feasible for calculation of distance scores (see below). All other manipulations with the structures including extraction of amino acid sequences and structure-based multiple sequence alignments were done using SwissPDBViewer (Guex and Peitsch 1997).

Distance Measure

The pairwise distance scores (listed in Table 2) were calculated according to the method of Johnson et al. (1990), with detailed methodology provided therein. Briefly, the distance scores (*D*) were based on the RMS and the PFTE (pairwise fractional topological equivalence: the number of pairs of equivalent α-carbon atoms used for calculation of the RMS value divided by the total number of α-carbons in the smaller of two compared structures). The RMS distances were converted to a similarity measure, SRMS, scaled between 0 and 1.0:

$$\mathrm{SRMS} = (1 - \mathrm{RMS})/\mathrm{cutoff\ rate} \quad (1)$$

Originally, the topological equivalence measure was introduced to facilitate discrimination between the pairs of protein folds that are strongly diverged, but cognate in their entirety, and the pairs that have common substructures, but dissimilar peripheral elements (Johnson et al. 1990). The distance score (*D*) was calculated as a weighted function

Table 1. Protein structures compared in this study

PDB code	Resolution (Å)	Description of the functional unit	DNA recognition and cleavage pattern (only 5′ strand is shown)	Reference
leri	2.7	*Eco*RI type II ENase, dimer	G$^{\downarrow}$AATTC[a]	Kim et al. (1990)
lrva	2.0	*Eco*RI type II ENase, dimer	GAT$^{\downarrow}$ATC[a]	Winkler et al. (1993)
lbhm	2.2	*Bam*HI type II ENase, dimer	G$^{\downarrow}$GATCC[a]	Newman et al. (1994)
lpvi	2.8	*Pvu*II type II ENase, dimer	CAG$^{\downarrow}$CTG[a]	Cheng et al. (1995)
lcfr	2.15	*Cfr*10I type II ENase, dimer	R$^{\downarrow}$CCGGY[a] (R = A or G, Y = T or C)	Bozic et al. (1996)
lfok (aa 391–568)	2.8	*Fok*I type IIs ENase, bipartite protein; catalytic domains dimerize during cleavage	GGATG(N)$_9$$^{\downarrow}$NNNN; recognizes a specific sequence and cleaves DNA nonspecifically at 9 bp away on the 5′ strand and 13 bp away on the 3′ strand	Wah et al. (1997)
lazo	1.7	MutH repair ENase, monomer	$^{\downarrow}$GATC; cleaves one strand	Ban and Yang (1998)
lavq	2.4	Phage λ exonuclease, trimer	Binds to a free end of dsDNA and degrades one strand in the 5′-to-3′ direction	Kovall and Matthews (1998)
NA[b]	2.2	*Bg*/I type II ENase, dimer	GCCNNNN$^{\downarrow}$NGGC[a]	Newman et al. (1998)

[a] $^{\downarrow}$ indicates the site of phosphodiester bond cleavage. All type II ENases cut their palindromic target identically in both strands.
[b] Not available.

Table 2. Structural distance matrix[a]

ENase	α-C	*Eco*RI	*Eco*RV	*Bam*HI	*Pvu*II	*Cfr*10I	*Fok*I	*Mut*H	λ-Exo	*Bgl*I
*Eco*RI	261	*	2.05/61	2.02/72	2.27/50	2.00/88	2.32/75	2.08/39	2.47/60	2.23/64
*Eco*RV	244	**1.1888**	*	2.14/71	2.18/67	2.42/76	2.46/68	2.41/90	2.00/68	2.21/116
*Bam*HI	208	**0.9791**	**1.0255**	*	2.41/48	1.95/81	2.10/67	2.24/41	2.20/51	2.50/70
*Pvu*II	156	**1.1059**	**0.8955**	**1.1749**	*	2.84/62	2.63/72	2.16/75	1.74/61	1.87/62
*Cfr*10I	283	**0.9884**	**1.1693**	**0.8875**	**1.0899**	*	1.93/88	2.04/48	2.03/65	2.27/79
*Fok*I	178	**0.9431**	**1.0407**	**0.9530**	**0.9521**	**0.7491**	*	2.38/45	1.92/63	2.10/75
*Mut*H	212	**1.3889**	**0.9599**	**1.4168**	**0.8244**	**1.2455**	**1.3017**	*	2.09/48	2.06/69
λ-*Exo*	226	**1.2962**	**1.0571**	**1.2583**	**0.8186**	**1.0952**	**0.9334**	**1.2660**	*	2.14/65
*Bgl*I	298	**1.2694**	**0.8432**	**1.1371**	**0.8511**	**1.1975**	**0.8853**	**1.0296**	**1.1339**	*

[a] The total number of α-carbon atoms in each structure is indicated. The cells above the diagonal show the RMS and the number of equivalent α-carbon atom pairs for the compared protein structures. The cells below the diagonal (in boldface) show the structure-based distances (*D*) calculated according to formula (2). A 3.5-Å cutoff value was used; the SRMS and the PFTE values were calculated based on the RMS and the number of equivalent α-C atoms divided by the total number of α-C atoms in the smaller structure, respectively, as indicated.

of the fractional topological equivalence (PFTE) and the scaled RMS distance (SRMS):

$$D = \ln\{\mathrm{PFTE}[(1 - \mathrm{PFTE}) + (1 - \mathrm{SRMS})]/2 + \mathrm{SRMS}(\mathrm{PFTE} + \mathrm{SRMS})/2\} \quad (2)$$

Following Johnson et al. (1990), the inverse weights were introduced to moderate the influence of the PFTE at small distances, where the RMS distance provides a better estimation of similarity, and to reduce the contribution of the RMS at larger distances, where it is calculated based on smaller subsets of atoms.

Molecular Phylogeny

The structure-based phylogenetic tree of the ENase-like superfamily was inferred using the FITCH program from the PHYLIP package (Felsenstein 1993) based on the method of Fitch and Margoliash (1967). The distances listed in Table 2 were used. The validity of the presented phylogram was also tested using the neighbor-joining approach (Saitou and Nei 1987) for the same data set or narrowed by the jackknife method. Alternative coordinate sets from the PDB were also used and calculations were done for cutoff rates between 3.0 and 3.8 Å.

The multiple sequence alignment derived from structure superposition (available from the author upon request) served to infer "conventional" trees using a variety of substitution models and methods from the PHYLIP package (Felsenstein 1993). Branch stability was evaluated using the bootstrap test with 1000 replications.

For both structure- and sequence-based tree inference the data input order was randomized, and the "global rearrangements" option was applied, where only possible.

The tree was drawn using the TreeView program (Page 1996).

Results and Discussion

Nine structures of restriction ENases and proteins reported to possess similar three-dimensional folds and catalytic mechanisms (Table 1) were aligned based on the three-dimensional position of their α-carbon atoms. Results obtained from two pairwise superposition methods (Zhang and Ding 1993; Holm and Sander 1997) run with different parameters showed absolute agreement. All conserved active-site residues and β-strands of the structural core were perfectly matched, judged from attempts to improve the three-dimensional alignment by a

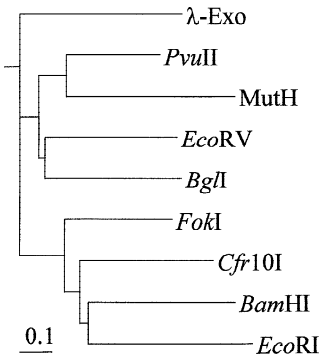

Fig. 1. The structure-based phylogenetic tree of the ENase-like fold superfamily inferred using the Fitch and Margoliash (1967) method. Restriction ENases are shown in *italics*. The *scale* below the tree indicates 0.1 *D* (the distance measure).

semiautomatic option of SwissPDB Viewer (Guex and Peitsch 1997).

To infer the phylogenetic tree of the ENase-like superfamily based on aligned protein structures (Fig. 1), the distance matrix method of Fitch and Margoliash (1967) was applied. Calculations were based on distance values [introduced by formula (2)], reflecting the scaled RMS and PFTE distances between compared structures, as described by Johnson et al. (1990). Neighbor-joining analysis (Saitou and Nei 1987) of the same data set or narrowing of it by the jack knife method yielded identical subtopologies. The validity of the presented phylogram was also tested using alternative coordinate sets from the PDB and changing the cutoff rate between 3.0 and 3.8 Å. All topologies obtained were identical, with negligible differences in branch lengths (data not shown).

I also attempted to infer an evolutionary tree from a structure alignment-based multiple sequence alignment using a variety of methods with different parameters to enable comparison of "conventional" and structure-based phylograms. However, none of the algorithms could produce even moderately robust trees and all the sequence-based trees obtained differed markedly in terms of both topology and branch lengths (data not shown). This implies the suitability of the structure-based approach in the study of molecular evolution of extremely diverged protein families like ENases and related enzymes.

Figure 1 shows the tree, which exhibits that the blunt end-generating ENases *Pvu*II and *Eco*RV (with their close homologues MutH and *Bgl*I ENases) and those leaving 5′ four-base overhangs form two separate clades. It confirms previous reports of the highest degree of divergence between these two classes of type II ENases (Winkler et al. 1993; Newman et al. 1994; Bozic et al. 1996). But the phylogenetic analysis reveals many more details. The mismatch repair ENase MutH must have diverged after the radiation of the two main lineages. Apparently, the most parsimonious hypothesis would be that MutH repair functions evolved from a restriction-like enzyme, rather than that the restriction function appeared independently in both main branches of the ENase tree. Furthermore, evidence for branching out the *Fok*I catalytic domain [which alone confers virtually no sequence specificity (Bitinaite et al. 1998)] at the bottom of the other branch suggests that the "5′ four-base overhand cutters" could have diverged from a nonspecific ancestor. It is possible that the *Fok*I lineage (e.g., type IIs ENases) retained the dimerization and DNA-binding mode but lost its sequence specificity after the fusion of a catalytic domain with a separate target recognizing domain. Although the data presented do not support the hypothesis of its direct relatedness to the "ancient ENase," it appears that the nonspecific *Fok*I cleavage domain is evolutionary older than other enzymes of its branch. It is reasonable that *Cfr*10I, exhibiting "relaxed" sequence specificity (Bozic et al. 1996), diverged from an evolutionary intermediate between *Fok*I and the highly specific *Bam*HI and *Eco*RI. It is worth emphasizing that inherently nonspecific cleavage domains of type I and type III RM systems, whose DNA-binding domains are not homologous to that of *Fok*I, displayed a much higher degree of similarity with the homologues of *Fok*I than with the *Eco*RV- and *Pvu*II-related enzymes, as assessed by threading methods, used for quantitative evaluation of sequence/structure compatibility (J.M. Bujnicki and L. Rychlewski, to be published).

It is probable that the λ-exo (and related proteins, for which, however, structural data are unavailable) branched out prior to the divergence of "endonucleolytic" subfamilies. Such speciation must have been more ancient than the burst of endonucleases presently comprising mainly prokaryotic RM systems, spread widely by lateral gene transfer (Jeltsch and Pingoud 1996). It would be interesting to determine whether any eukaryotic proteins also possess domains related to the ENase-like superfamily and, if this is the case, whether they originate from the Cenancestor or have been acquired by horizontal transfer. The answer to these questions coupled with similar studies on modification methylases would help to explain the origin of RM systems and their basic biological role. Large-scale structure prediction, three-dimensional threading, and homology modeling experiments aimed at solving that problem have been initiated recently (J.M. Bujnicki and L. Rychlewski, unpublished data).

It is tempting to speculate that the quantitative estimation of the possible relatedness between homologous proteins should correlate with the relatedness of the corresponding recognition sequences. However, with the data currently available it appears that most dramatic changes in the cleavage pattern are achieved not by variations in the sequence and structure of a monomer but, rather, by alterations in the mode of dimerization

and topology of the respective protein–DNA complexes (Bozic et al. 1996; Newman et al. 1998). Therefore, the similarity between target sequences cannot be considered as a measure of evolutionary distance between ENases. This demonstrates the superiority of the approach presented above over the combined geno/phenotypic methodology of Jeltsch et al. (1995) in quantitative analysis of relatedness among ENases. The topology of the tree and branching pattern is fully confirmed by functional and structural considerations reported for each structure at the time of its determination. Hence, the results presented here allow for integration of the most important aspects of incomplete comparative studies published to date.

A major challenge for future research is to predict the details of new ENase–DNA complexes by means of structure prediction and homology modeling. Such insights may provide a crucial key to the elusive goal of engineering the specificity of restriction enzymes. However, ENases still remain beyond the reach of conventional tertiary structure prediction methods because of the general lack of success in sequence similarity database searches. The evolutionary tree presented is important not only for the classification of ENases or choice of possible modeling templates, but also as a guide for large-scale mutagenesis experiments, including construction of hybrid proteins with novel specificities. Further structural characterization of additional nucleases should provide necessary validation of the presented evolutionary model. The anticipated structure solution of type I and type III ENases may contribute significantly to our understanding of the sequence–structure–function relationship and the evolution of DNA sequence specificity in the phylogenetic history of the ENase catalytic domain.

Acknowledgments. I would like to thank Prof. Simon E. Phillips for providing coordinates of *Bgl*I and Dr. Baohong Zhang for help with his 3D-alignment server. I thank also Dr. Monika Radlinska and Wilbur Tong for critical reading of the manuscript.

References

Aggarwal AK (1995) Structure and function of restriction endonucleases. Curr Opin Struct Biol 5:11–19

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

Bernstein FC, Koetzle TF, Williams GJ, et al. (1977) The Protein Data Bank: A computer-based archival file for macromolecular structures. J Mol Biol 112:535–542

Bickle TA, Kruger DH (1993) Biology of DNA restriction. Microbiol Rev 57:434–450

Bitinaite J, Wah DA, Aggarwal AK, Schildkraut I (1998) *Fok*I dimerization is required for DNA cleavage. Proc Natl Acad Sci USA 95:10570–10575

Bozic D, Grazulis S, Siksnys V, Huber R (1996) Crystal structure of *Citrobacter freundii* restriction endonuclease *Cfr*10I at 2.15 Å resolution. J Mol Biol 255:176–186

Bujnicki JM (1999) Comparison of protein structures reveals monophyletic origin of AdoMet-dependent methyltransferase family and mechanistic convergence rather than recent differentiation of N4-cytosine and N6-adenine DNA methylation. In Silico Biol 1:16 (http://www.bioinfo.de/isb/1999/01/0016/)

Cheng X, Balendiran K, Schildkraut I, Anderson JE (1995) Crystal structure of the *Pvu*II restriction endonuclease. Gene 157:139–140

Felsenstein (1989) PHYLIP (phylogeny inference package), version 3.5. Cladistics 5:164–166 (http://evolution.genetics.washington.edu/phylip.html)

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279–284

Fukami-Kobayashi K, Teteno Y, Nishikawa K (1999) Domain dislocation: A change of core structure in periplasmic binding proteins in their evolutionary history. J Mol Biol 286:279–290

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 18:2714–2723

Holm L, Sander C (1997) Dali/FSSP classification of three-dimensional protein folds. Nucleic Acids Res 25:231–234

Horton NC, Newberry KJ, Perona JJ (1998) Metal ion-mediated substrate-assisted catalysis in type II restriction endonucleases. Proc Natl Acad Sci USA 95:13489–13494

Jeltsch A, Pingoud A (1996) Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. J Mol Evol 42:91–96

Jeltsch A, Alves J, Maass G, Pingoud A (1992) On the catalytic mechanism of *Eco*RI and *Eco*EV. A detailed proposal based on biochemical results, structural data and molecular modeling. FEBS Lett 304: 4–8

Jeltsch A, Kroger M, Pingoud A (1995) Evidence for an evolutionary relationship among type-II restriction endonucleases. Gene 160:7–16

Johnson MS, Sutcliffe MJ, Blundell TL (1990) Molecular anatomy: Phyletic relationships derived from three-dimensional structures of proteins. J Mol Evol 30:43–59

Kim Y, Grable JC, Love R, Green PJ, Rosenberg JM (1990) Refinement of *Eco*RI endonuclease crystal structure: A revised protein chain tracing. Science 249:1307–1309

Kovall RA, Matthews BW (1998) Structural, functional, and evolutionary relationships between λ-exonuclease and the type II restriction endonucleases. Proc Natl Acad Sci USA 95:7893–7897

Kroger M, Hobom G, Schutte H, Mayer H (1984) Eight new restriction endonucleases from *Herpetosiphon giganteus*—Divergent evolution in a family of enzymes. Nucleic Acids Res 12:3127–3141

Newman M, Strzelecka T, Dorner LF, Schildkraut I, Aggarwal AK (1994) Structure of restriction endonuclease *Bam*HI and its relationship to *Eco*RI. Nature 368:660–664

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

Page RDM (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. CABIOS 12:357–358 (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html)

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

Roberts RJ, Macelis D (1999) REBASE-restriction enzymes and methylases. Nucleic Acids Res 27:312–313

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Siksnys V, Timinskas A, Klimasauskas S, Butkus V, Janulaitis A (1995) Sequence similarity among type-II restriction endonucleases, related by their recognized 6-bp target and tetranucleotide-overhang cleavage. Gene 157:311–314

Skirgaila R, Grazulis S, Bozic D, Huber R, Siksnys V (1998) Structure-based redesign of the catalytic/metal binding site of *Cfr*10I restriction endonuclease reveals importance of spatial rather than se-

quence conservation of active center residues. J Mol Biol 279:473–481

Venclovas C, Timinskas A, Siksnys V (1994) Five-stranded beta-sheet sandwiched with two alpha-helices: A structural link between restriction endonucleases *Eco*RI and *Eco*RV. Proteins 20:279–282

Wah DA, Hirsch JA, Dorner LF, Schildkraut I, Aggarwal AK (1997) Structure of the multimodular endonuclease *Fok*I bound to DNA. Nature 388:97–100

Winkler FK, Banner DW, Oefner C, et al. (1993) The crystal structure of *Eco*RV endonuclease and of its complexes with cognate and non-cognate DNA fragments. EMBO J 12:1781–1795

Withers BE, Ambroso LA, Dunbar JC (1992) Structure and evolution of the *Xcy*I restriction-modification system. Nucleic Acids Res 20: 6267–6273

Zhang B, Ding D (1993) A new method for pairwise comparison of protein structures. Acta Biophys Sin 9:353–361