Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization and preliminary diffraction analysis of the *Hinc*II restriction endonuclease–DNA complex

Nancy C. Horton,[a] Lydia F. Dorner,[b] Ira Schildkraut[b] and John J. Perona[a]*

[a]Department of Chemistry and Interdepartmental Program in Biochemistry and Molecular Biology, University of California at Santa Barbara, Santa Barbara CA 93106-9510, USA, and [b]New England Biolabs, 32 Tozer Road, Beverly, MA 01915, USA

Correspondence e-mail: perona@chem.ucsb.edu

Received 25 May 1999
Accepted 2 August 1999

Crystals of the 60 kDa dimeric *Hinc*II restriction enzyme bound to a 12 base-pair dyad-symmetric duplex DNA carrying the specific 5′-GTCGAC recognition site have been obtained. Crystals grew by hanging-drop vapor diffusion from solutions containing polyethylene glycol 4000 as precipitating agent. The rod-shaped crystals belong to space group $I222$ (or $I2_12_12_1$), with unit-cell dimensions a = 66.9, b = 176.7, c = 256.0 Å. There are most likely to be two dimeric complexes in the asymmetric unit. A complete native data set has been collected from a high-energy synchrotron source to a resolution of 2.5 Å at 100 K, with an $R_{\rm merge}$ of 4.8%.

1. Introduction

Type II restriction endonucleases are superb models for the study of protein–DNA interactions because of their exceptionally high specificities for cleavage of target DNA sites (Roberts & Halford, 1993). Typically, mutation of a single nucleotide of a 4–8 base-pair dyad-symmetric recognition sequence decreases rate constants by 10^5-fold to 10^6-fold (Lesser *et al.*, 1990), as required for function in a bacterial host defensive system. Crystal structures of six type II restriction endonucleases have been determined: *Eco*RI (Rosenberg, 1991), *Eco*RV (Winkler *et al.*, 1993), *Pvu*II (Athanasiadis *et al.*, 1994; Cheng *et al.*, 1994), *Bam*HI (Newman *et al.*, 1995), *Cfr*10I (Bozic *et al.*, 1996) and *Bgl*I (Newman *et al.*, 1998). The type IIs enzyme *Fok*I has also been solved (Wah *et al.*, 1997). All of the structures except *Cfr*10I are determined bound to DNA. Little primary sequence similarity exists among the enzymes; thus, the use of sequence information to help understand structure–function relationships has been limited.

The four enzymes *Eco*RI, *Bam*HI, *Eco*RV and *Pvu*II are divisible into those which cleave leaving four-base 5′-overhangs (*Eco*RI, *Bam*HI) and those which cleave at the center step to produce blunt-ended products (*Eco*RV, *Pvu*II). While topologically similar folds characterize each subclass (Aggarwal, 1995), large differences in structure still remain. For example, comparison of *Eco*RV and *Pvu*II shows that while three central β-strands are superimposable with an r.m.s. deviation of 0.8 Å, flanking elements in the conserved active-site domain deviate substantially. The dimer interfaces of the two enzymes are completely unrelated. Most significantly, even the general modes of DNA recognition are not the same: *Pvu*II binds unbent B-form DNA using antiparallel β-structure, whereas *Eco*RV employs surface loops and bends the DNA sharply into the major groove.

The small number of known type II restriction-enzyme structures poses severe limitations on our ability to rationally engineer new sequence specificities. It is already known that the introduction of small numbers of amino-acid substitutions in the DNA-binding clefts of *Eco*RI and *Eco*RV does not result in specificity modification, but rather in reduced activities toward both wild-type and modified target sites (Heitman & Model, 1990; Flores *et al.*, 1995; Wenz *et al.*, 1994; Lanio *et al.*, 1996). An alternative approach toward understanding and ultimately manipulating sequence preferences is to generate a database of structures within a subclass where there is precedent for topological similarity within the context of different specificities. Comparisons among such enzymes should lead to an improved understanding of the roles of individual polypeptide regions and amino acids in generating the extraordinary DNA-sequence selectivities.

Here, we report crystals of restriction endonuclease *Hinc*II bound to a duplex DNA molecule containing the specific recognition site. *Hinc*II recognizes the sequence 5′-GTPyPuAC and cleaves it in a blunt-ended fashion at the center Py–Pu step (Kelly & Smith, 1970). It is thus identical with the blunt-cutting *Eco*RV (5′-GATATC) at the outer pair, with similar but degenerate specificity at the inner base-pair step. Interestingly, a limited sequence similarity of *Hinc*II with *Eco*RV in the region of the major-groove binding recognition loops (R-loops) of *Eco*RV has been noted (Thielking *et al.*, 1991). This similarity provides some further confidence that the *Hinc*II and *Eco*RV active-site domains will be topologically similar. It is known that *Eco*RV utilizes the sequence-dependent free-energy

© 1999 International Union of Crystallography
Printed in Denmark – all rights reserved

Table 1
DNA sequences used in cocrystallization.

1	CGGTCGACCG
2	CCGGTCGACCGG
3	ACGGTCGACCG
4	GCGGTCGACCG
5	ACGGTTAACCG
6	GCGGTTAACCG
7	ACCGGTCGACCGG
8	GCCGGTCGACCGG
9	ACCGGTTAACCGG
10	GCCGGTTAACCGG

Table 2
Native *Hinc*II–DNA data-collection statistics.

Resolution (Å)	$R_{\rm merge}$	$\langle I\rangle/\sigma$	Total number of observations	Number of unique reflections	Completeness (%)	Multiplicity
11.18	0.040	44.2	1239	509	79.4	2.4
7.91	0.029	41.6	2477	999	89.2	2.5
6.45	0.034	33.1	3304	1315	92.0	2.5
5.59	0.030	30.2	3927	1572	93.8	2.5
5.00	0.030	29.5	4414	1807	95.2	2.4
4.56	0.030	31.6	4869	2001	96.1	2.4
4.23	0.030	30.1	5301	2196	97.5	2.4
3.95	0.035	26.9	5608	2342	96.9	2.4
3.73	0.040	24.1	5867	2502	97.5	2.3
3.54	0.047	20.2	5874	2588	95.5	2.3
3.37	0.051	18.1	5850	2681	94.0	2.2
3.23	0.060	15.2	5669	2702	90.7	2.1
3.10	0.075	12.3	5619	2746	88.5	2.0
2.99	0.089	9.9	5467	2756	85.5	2.0
2.89	0.110	8.0	5425	2767	83.1	2.0
2.80	0.119	7.2	5318	2792	81.0	1.9
2.71	0.140	5.8	5317	2819	79.4	1.9
2.64	0.168	4.9	5166	2797	76.5	1.8
2.56	0.191	4.2	5182	2810	75.0	1.8
2.50	0.223	3.6	5062	2815	73.0	1.8
Overall	0.047	16.6	96955	45516	86.6	2.1

cost of base-pair unstacking to generate part of the specificity for TA at the center step, where the DNA is bent sharply (Martin *et al.*, 1999). Some further insight into this interesting discrimination process is expected from the structure of *Hinc*II, which can accommodate both TA and CG at the center base-pair step of its recognition site.

2. Purification, crystallization and diffraction data collection

*Hinc*II endonuclease was overexpressed in *Escherichia coli* as described (Rees *et al.*, 1988). A 1 l overnight culture was grown in rich medium at 310 K and used as the inoculum for a 100 l fermentation. Induction of expression at a corrected Klett value of 120–150 was by addition of 0.5 m*M* IPTG. Cells were harvested and stored at 203 K prior to thawing and lysis of 125 g of cells in a Gaulin cell disruptor. The enzyme was purified through five successive column chromatographic steps on DEAE Sepharose (Pharmacia), phosphocellulose (Whatman), heparin sepharose FF (Pharmacia), Mono Q (HR 10/10; Pharmacia) and G-75 Sephadex. The enzyme was assayed throughout *via* cleavage of λ DNA. After the final column the purified *Hinc*II endonuclease was dialyzed into a storage buffer containing 0.1 *M* NaCl, 10 m*M* Tris–HCl (pH 7.4), 0.1 m*M* EDTA, 1 m*M* DTT, 50% glycerol and frozen in aliquots at 203 K.

For crystallization trials, *Hinc*II stored at 203 K was thawed on ice and dialyzed into a buffer containing 10 m*M* HEPES (pH 7.5), 150 m*M* NaCl, 1 m*M* DTT, 0.1 m*M* EDTA and then concentrated to 8 mg ml^{-1} in a Microcon concentrator (Amicon). Glycerol was added to 50% volume and the enzyme was stored at 253 K until just prior to use. Aliquots were then dialyzed into the concentration buffer to remove the glycerol and reconcentrated to at least 8 mg ml^{-1}. Self-complementary oligodeoxynucleotides shown in Table 1 were prepared synthetically by standard techniques and purified by reverse-phase HPLC (Aggarwal, 1990). The DNA was lyophilized and resuspended at high concentration in water. *Hinc*II/DNA solutions were prepared to give final protein concentration of 6–8 mg ml^{-1} and a 1.5-fold molar excess of DNA. Crystallization of *Hinc*II–DNA complexes was performed by hanging-drop vapor diffusion in Linbro plates using siliconized glass cover slips. Crystallization conditions were screened at 290 K using varying concentrations of polyethylene glycol 4000 (PEG 4K) and NaCl, with 0.1 *M* buffers ranging in pH from 4.5 to 8.5. Crystals of approximate dimensions 0.1 mm or larger were screened for diffraction after mounting in glass capillaries, using a Rigaku RU-200 rotating-anode X-ray source at ambient temperatures. Each of the ten DNA sequences screened produced crystalline material. Four of these crystals (from DNA sequences 2, 6, 9 and 10; Table 1) diffracted weakly to 5–10 Å resolution. However, the sequence 5′-GCCG-GTCGACCGG (sequence 8), possessing a 5′-G overhang, yielded measurable diffraction amplitudes to 3.5 Å resolution on the rotating-anode X-ray source when co-crystallized with *Hinc*II. Gel electrophoresis of carefully washed and dissolved crystals demonstrated the presence of both protein and DNA (data not shown). These crystals were grown at 290 K by mixing 1 µl of *Hinc*II–DNA complex (7.2 mg ml^{-1}) with 1 µl of crystallization solution containing 20% PEG 4K, 0.1 *M* sodium citrate (pH 5.5) and 0.12 *M* NaCl.

Crystals of this *Hinc*II–DNA complex were exchanged into a cryoprotectant consisting of 25% PEG 4K, 0.3 *M* NaCl, 0.1 *M* sodium citrate (pH 5.5) and 30% glycerol. The crystal was flash-frozen in a stream of nitrogen gas held at 100 K at Stanford Synchrotron Radiation Laboratory beamline 1-5. A native data set (Table 2) was collected on a CCD detector from 1° oscillation photographs spanning two 30° wedges of reciprocal space at a crystal-to-detector distance of 220 mm. The exposure time per image was 4 min. Data were processed with *MOSFLM* (Leslie, 1996) and *SCALA* (Evans, 1997). These data are 86.6% complete to 2.5 Å resolution; completeness is somewhat lower at higher resolution owing to the square detector surface. The $I/\sigma(I)$ value in the highest resolution shell is 3.6. The calculated Matthews coefficient of 2.78 Å^3 Da^{-1} suggests two enzyme dimers/DNA duplexes per asymmetric unit, although the presence of either three or five enzyme monomers/DNA single strands would also be within the acceptable range of solvent content. Various strategies for phasing including iodination or bromination of the DNA, conventional heavy-metal soaks and crystallization of selenomethionyl-substituted enzyme are being employed.

References

Aggarwal, A. K. (1990). *Methods*, **1**, 83–90.
Aggarwal, A. K. (1995). *Curr. Opin. Struct. Biol.* **5**, 11–19.
Athanasiadis, A., Vlassi, M., Kotsifaki, D., Tucker, P. A., Wilson, K. S. & Kokkinidis, M. (1994). *Nature Struct. Biol.* **1**, 469–475.
Bozic, D., Grazulis, S., Siksnys, V. & Huber, R. (1996). *J. Mol. Biol.* **255**, 176–186.
Cheng, X., Balendiran, K., Schildkraut, I. & Anderson, J. E. (1994). *EMBO J.* **17**, 3927–3935.

Evans, P. R. (1997). *Jnt CCP4/ESF–EACBM Newslett. Protein Crystallogr.* **33**, 22–24.
Flores, H., Osuna, J., Heitman, J. & Soberon, X. (1995). *Gene*, **157**, 295–301.
Heitman, J. & Model, P. (1990). *EMBO J.* **9**, 3369–3378.
Kelly, T. J. & Smith, H. O. (1970). *J. Mol. Biol.* **51**, 393–409.
Lanio, T., Selent, U., Wenz, C., Wende, W., Schulz, A., Adiraj, M., Katti, S. B. & Pingoud, A. (1996). *Protein Eng.* **9**, 1005–1010.
Leslie, A. G. W. (1996). *Jnt CCP4/ESF–EACBM Newslett. Protein Crystallogr.* **32**, 7–8.
Lesser, D. R., Kurpiewski, M. R. & Jen-Jacobson, L. (1990). *Science*, **250**, 776–786.
Martin, A. M., Sam, M. D., Reich, N. O. & Perona, J. J. (1999). *Nature Struct. Biol.* **6**, 269–277.
Newman, M., Lunnen, K., Wilson, G., Greci, J., Schildkraut, I. & Phillips, S. E. V. (1998). *EMBO J.* **17**, 5466.
Newman, M., Strzelecka, T., Dorner, L. F., Schildkraut, I. & Aggarwal, A. K. (1995). *Science*, **269**, 656–663.
Rees, P. A., Nwankwo, G., Wilson, G. G. & Benner, J. S. (1988). *Gene*, **74**, 37.
Roberts, R. J. & Halford, S. E. (1993). *Nucleases*, edited by S. M. Linn, R. S. Lloyd & R. J. Roberts, 2nd ed., pp. 35–88. Cold Spring Harbor, USA: Cold Spring Harbor Laboratory Press.
Rosenberg, J. (1991). *Curr. Opin. Struct. Biol.* **1**, 104–113.
Thielking, V., Selent, U., Kohler, E., Wolfes, H., Pieper, U., Geiger, R., Urbanke, C., Winkler, F. K. & Pingoud, A. (1991). *Biochemistry*, **30**, 6416–6422.
Wah, D. A., Hirsch, J. A., Dorner, L. F., Schildkraut, I. & Aggarwal, A. K. (1997). *Nature (London)*, **388**, 97–100.
Wenz, C., Selent, U., Wende, W., Jeltsch, A., Wolfes, H. & Pingoud, A. (1994). *Biochem. Biophys. Acta*, **1219**, 73–80.
Winkler, F. K., Banner, D. W., Oefner, C., Tsernoglou, D., Brown, R. S., Heathman, S. P., Bryan, R. K., Martin, P. D., Petratos, K. & Wilson, K. S. (1993). *EMBO J.* **12**, 1781–1795.