

Refinement of Eco RI Endonuclease Crystal Structure: A Revised Protein Chain Tracing



Youngchang Kim; John C. Grable; Robert Love; Patricia J. Greene; John M. Rosenberg

Science, New Series, Vol. 249, No. 4974. (Sep. 14, 1990), pp. 1307-1309.

Stable URL:

<http://links.jstor.org/sici?sici=0036-8075%2819900914%293%3A249%3A4974%3C1307%3AROEREC%3E2.0.CO%3B2-6>

Science is currently published by American Association for the Advancement of Science.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/aaas.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

Refinement of Eco RI Endonuclease Crystal Structure: A Revised Protein Chain Tracing

The original model of Eco RI endonuclease (1) proved refractory to crystallographic refinement, whereby model coordinates were adjusted to optimize the fit to the observed diffraction data (the best *R* factor for the original model is 0.25). We have now obtained five additional isomorphous derivatives by cocrystallizing the protein with the halogenated oligonucleotides: U¹CGC-GAATTCGCG, TCGCGAAU¹TCGCG, TC-GCGAATU¹CGCG, TCGCGAATTCGC¹G and U^{Br}CGCGAAU^{Br}TCGCG^{Br}G (2) (Eco RI site underlined). The heavy atoms were located by difference Patterson and Fourier methods (with the use of phases from the original platinum derivative); halogen atoms were found at the locations expected from the earlier structure (a minor exception is the first derivative, which suggests that the unpaired thymidine may be in the *syn* conformation). The phasing power for the platinum derivative is 2.09; those for the halogen derivatives are 1.76, 2.11, 1.75, 1.67, and 1.96, respectively. The mean figure of merit is 0.59 for all six derivatives to 2.8 Å resolution.

The new data yielded a multiple isomorphous replacement (MIR) electron density map that suggests a different connectivity between elements of secondary structure in the protein (3). A model based on that connectivity has refined robustly by the X-PLOR and Konnert-Hendrickson methods (4) to an *R* factor of 0.20 (no ordered solvent has been included in this model, nor have coordinates for the first 16 amino acid residues, which appear disordered) (5). The root-mean-square deviations in bond distance, angle distance, and 1-4 dihedral distance are 0.016 Å, 0.031 Å, and 0.031 Å, respectively, in the current cycle of Konnert-Hendrickson. The real space *R* factor (6), and Ramachandran plots are well behaved. The building block decomposition of Unger *et al.* is standard (7), an indication that the backbone geometry is consistent with other protein structures that have been well determined to high resolution. Coordinates for C_α atoms of the protein and all atoms (other than hydrogen) of the DNA have been deposited with the Brookhaven Data Bank.

We also cocrystallized the protein with a longer oligonucleotide [15 nucleotides (nt)],

with the sequence TCGTGGGAATTCACG (space group R32, *a* = *b* = 128.1 Å, *c* = 148.1 Å). We obtained a K₂PtCl₄ derivative (phasing power 1.76) and an iodine derivative by means of a synthetic oligonucleotide in which the first C is replaced by C^I (phasing power 1.55). Anomalous scattering data were collected for both derivatives; the anomalous phasing powers are 1.63 and 1.55, respectively. The average figure of merit is 0.62 to 3.5 Å resolution. The MIR electron density map of the 15-nt complex was analyzed independently from that of the 13-nt complex described above. A chain tracing was obtained for the protein within the 15-nt complex; this tracing is essentially identical to that obtained for the 13-nt complex (8). Final analysis of the 15-nt complex awaits further refinement.

The new chain tracing (Fig. 1) is based on some new elements of electron density and a new interpretation that alters the assignment of specific amino acid residues to some of the original features. Of particular note are (i) the extended chain motif, (ii) the β bridge, (iii) the placement of Glu¹¹¹, and (iv) a β-loop-α topological motif.

The extended chain motif (Met¹³⁷ to Ala¹⁴²) is a segment of extended polypeptide chain that runs through the major groove of the DNA, roughly parallel to the DNA backbone (Fig. 2). Although the φ-ψ torsion angles of this segment are in the β region of the Ramachandran plot, the conformation is not sufficiently regular to be an isolated strand of β sheet. The motif is anchored at one end by the inner recognition helix, while its other end is fixed by the inner arm. There are several contacts between residues in the extended chain and some of the bases within the Eco RI recognition site. These are in addition to contacts from the recognition helices (5). The segment of the gene that codes for this motif is a mutational hotspot (9), which is not surprising given its structure. Most of the contacts from the extended chain motif are made to pyrimidine bases, consistent with the suggestion of Heitman and Model that the protein interacts with pyrimidines (10) as well as with purines. An extended chain motif has not been seen in previously reported DNA-protein cocrystal structures.

The β bridge is part of the loop connecting β1 to β2, which includes an α helix (α3) and a pair of antiparallel β strands (βⁱ and βⁱⁱ) that are not part of the main β sheet (11); α3 and especially the segment from Ile⁷³ to Asn⁸⁵ pack primarily against the twofold-related subunit rather than their own subunit. Specifically, the β bridge is formed by βⁱ and βⁱⁱ because they span the gap between the globular units. βⁱⁱ is in close proximity to the DNA backbone (Fig. 3), forming one side of the cleft where the DNA backbone is bound.

The current model places Glu¹¹¹ near the scissile bond of the DNA. Mutations at this locus have been identified by Modrich and co-workers that reduce the cleavage activity of the enzyme but do not alter (or even enhance) DNA binding (12). It is interesting that Glu¹¹¹ is part of β3, specifically the carboxyl-terminal portion, which runs parallel to the DNA backbone in a manner similar to the β bridge.

The β-loop-α motif can be seen in the topological similarity of the loops that connect β3 to α4 and β4 to α5 (Fig. 1). These loops form the arms, which are segments of structure that extend from the central domains and nearly surround the DNA (Fig. 2). The inner arm is the loop connecting β3 to the extended chain motif; it contacts the DNA backbone. It is buttressed by the outer arm, which is the loop connecting β4 to α5, as in the previous model. This topological symmetry, which is at the heart of the nucleotide binding fold, suggests that β-loop-α may be a general topological motif for nucleic acid binding.

Some of our previous discussion (1) was based on direct observation of the electron density map, particularly the spatial relations of the major structural elements (double helix, α helices, and β sheets). These conclusions have not changed because they are independent of any interpretation that resulted in the assignment of amino acid residues to specific features of electron density. The central kink, which unwinds the DNA by approximately 25 degrees, is a prominent feature of all the electron density maps. The helical recognition motif, which consists of a parallel bundle of four α helices penetrating the major groove of DNA, is similarly unaffected. This motif is different from those of helix-turn-helix, zinc finger, and leucine zipper.

The current model retains the topology in which a five-stranded, largely parallel β sheet is surrounded by α helices (Fig. 1). The order and direction of the strands within the β sheet remain unchanged. The antiparallel motif (β1-β3) is still the foundation for the strand scission site, and the parallel motif (β3-β5) is likewise the foundation for

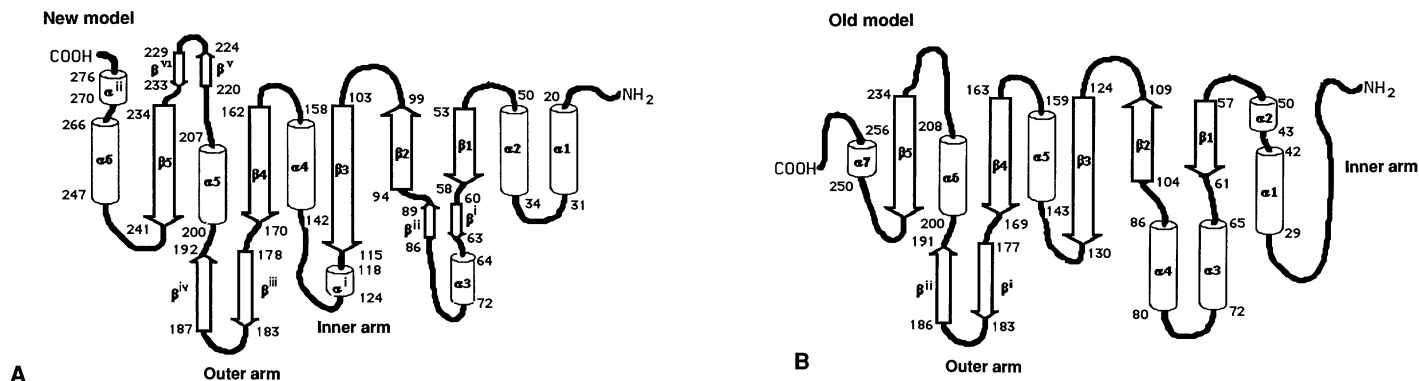


Fig. 1. (A) The topological structure of the new chain tracing for Eco RI endonuclease. Strands of β sheet are represented by arrows and α helices by cylinders. The strands of the major β sheet are shown with Arabic numerals sequentially from the amino terminus; strands in loop regions are indicated with sequential Roman numerals. Similarly, major α helices are shown with sequential Arabic numerals, while two short helices are indicated with Roman numerals. The amino acid residue numbers at the beginning and end of each element of secondary structure are indicated. (B) The old chain tracing is shown with similar labeling. (C) The three-dimensional flow of the polypeptide chain of one subunit of the enzyme is shown in stereo figures generated by Priestle's program (13). The second subunit, which is identical, is not shown for clarity. α helices are represented by coiled ribbons, strands of β sheets by arrows, and loops by narrow ribbons. The arm is toward the lower right portion of the figure, approximately in the plane of the figure. The β bridge (see text) and $\alpha 3$ projects out of the plane of the figure toward the viewer and is located to the center left. (D) A similar view including a rough approximation to the location of the DNA backbones, which are represented by β arrows.

the direct protein-DNA contacts. Helices $\alpha 4$ and $\alpha 5$ play a central role in the sequence specific interaction and are also referred to as the inner and outer recognition helices, respectively. The similarity between the parallel motif and nucleotide binding fold is thus enhanced.

In summary, new isomorphous derivative data suggest a revised chain tracing for the crystal structure Eco RI endonuclease that has refined satisfactorily. The central kink, which unwinds the DNA, is unaffected by the new data; likewise the major groove is penetrated by a parallel bundle of four α helices. Principal changes in the protein

include altered connectivity through loops and helices at the surface of the complex that nonetheless retains the original largely parallel five-stranded α - β topology. The new data reveal a β bridge that interacts with the DNA backbone, a repeated β -loop- α topological motif for nucleic acid recognition, and an extended chain motif that runs through the major groove and contacts some of the bases in the Eco RI recognition site.

*Present address: Agouron Pharmaceuticals, Inc., 11025 North Torrey Pines Road, La Jolla, CA 92037.
 †To whom correspondence should be addressed.

YOUNGCHANG KIM
 JOHN C. GRABLE
 ROBERT LOVE*
 Departments of Biological Sciences and
 Crystallography,
 University of Pittsburgh,
 Pittsburgh, PA 15260
 PATRICIA J. GREENE
 Department of Biochemistry and Biophysics,
 University of California,
 San Francisco, CA 94143
 JOHN M. ROSENBERG†
 Departments of Biological Sciences and
 Crystallography,
 University of Pittsburgh

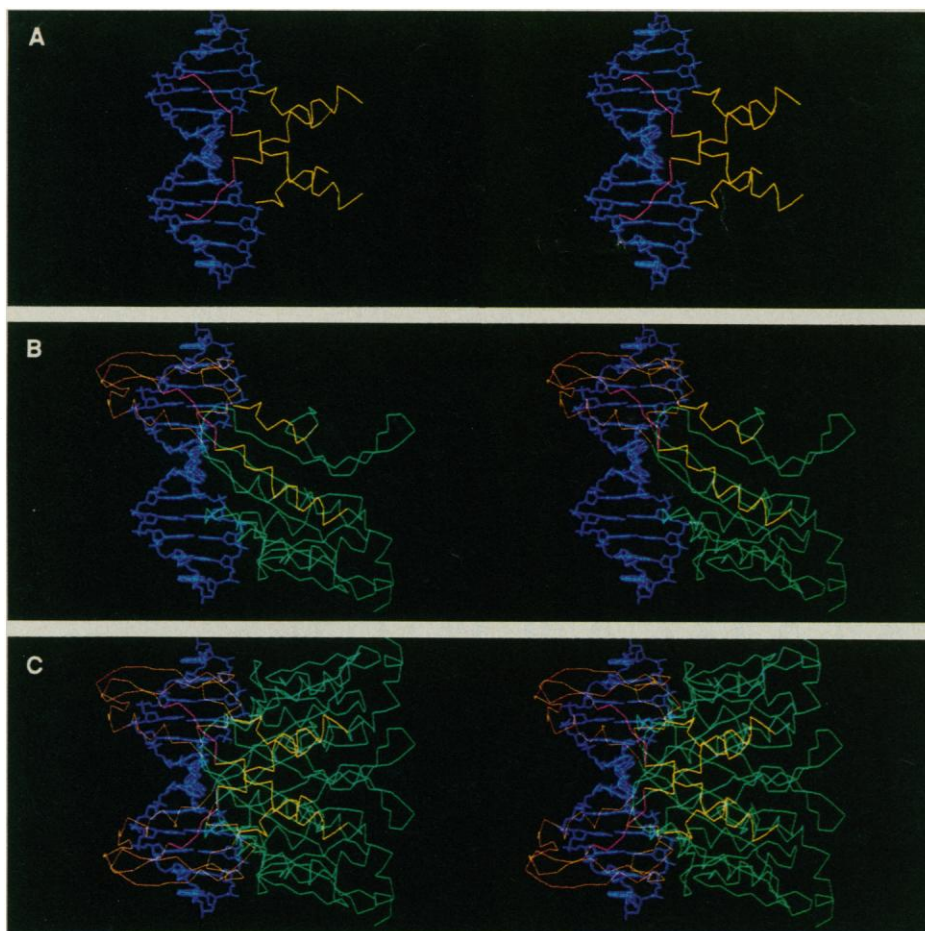


Fig. 2. (A) The C_{α} atoms from the segments of polypeptide chain that form the principle recognition elements are shown against the DNA. The four recognition helices, two from each subunit, are yellow, and the extended chain motifs from each subunit are pink. The molecular axis of twofold symmetry is horizontal, in the plane of the figure. (B) C_{α} atoms for one entire subunit of Eco RI endonuclease are shown against the DNA. The recognition helices and extended chain motif are yellow and pink, respectively, as in the previous panel. The arm is orange, and the rest of the polypeptide chain is green. (C) Both subunits of the protein dimer are shown with the DNA, colored as in the previous panel.

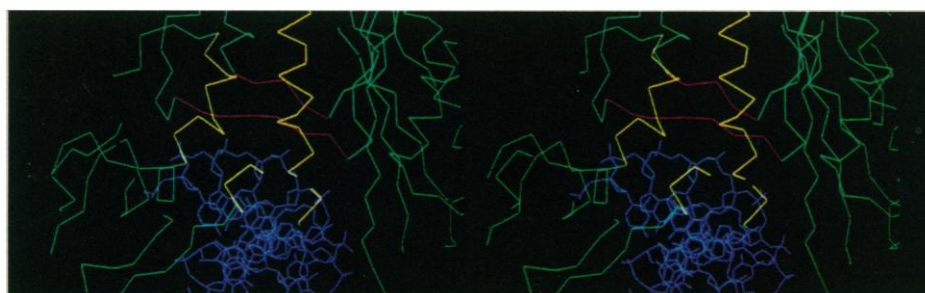


Fig. 3. A cross section of the complex showing the β bridge. The view is down the average DNA helix axis; only those atoms at the approximate depth of the β bridge are shown. The DNA is blue, the β bridge is red, the visible portions of the recognition helices are yellow, and the rest of the protein is green. β^{ii} is the red strand that is close to the DNA backbone. β^{3} can also be seen; it is approximately vertical and follows the DNA backbone in the lower right portion of the figure.

REFERENCES AND NOTES

1. J. A. McClarin, C. A. Frederick, B.-C. Wang, P. Greene, H. W. Boyer, J. Grable, J. M. Rosenberg, *Science* **234**, 1526 (1986).
2. U^{I} is 5-iododeoxyuracil, C^{I} is 5-iodocytosine, and U^{Br} and C^{Br} are the corresponding 5-bromopyrimidines. Some of the halogenated oligonucleotides contain base analogs in the recognition site, and further analyses of the structures of these derivatives are required to detect subtle changes that could affect DNA recognition.
3. Some but not all of the connectivity changes suggested by the new diffraction data were anticipated by electron density maps obtained with phases calculated from fragments of the original structure. Unfortunately, those maps were ambiguous and analysis of them proved inconclusive. Complete native data were re-collected to 2.7 Å resolution on a Nicolet arc detector as were data for the platinum derivative (1), but the resultant maps were equally inconclusive; hence the need to obtain additional isomorphous derivatives.
4. A. Brünger, J. Kuriyan, M. Karplus, *Science* **235**, 458 (1987); W. Hendrickson, *Methods Enzymol.* **115**, 252 (1985).
5. Y. Kim, P. J. Greene, J. M. Rosenberg, in preparation.
6. C.-I. Branden and T. A. Jones, *Nature* **343**, 687 (1990); Y. Kim, unpublished program.
7. R. Unger, D. Harel, S. Wherland, J. L. Sussman, *Proteins* **5**, 355 (1989).
8. J. Grable, Y. Kim, P. J. Green, J. M. Rosenberg, in preparation.
9. S. D. Yanofsky, R. Love, J. A. McClarin, J. M. Rosenberg, H. B. Boyer, P. J. Greene, *Proteins* **2**, 273 (1987).
10. J. Heitman and P. Model, *ibid.* **7**, 185 (1990).
11. We use the term loop in the sense of Richardson [J. S. Richardson, *Adv. Protein Chem.* **34**, 167 (1981)], which is the stretch of polypeptide chain that interconnects the strands of the principal β sheet or crossover helices, the latter being those α helices that pack against the β sheet and are also necessary for the β strands to connect in a parallel rather than antiparallel direction. By this definition loops often contain α helices or β hairpins (or both) that do not contribute to forming the core of the α - β domain.
12. K. King, D. Wright, P. Modrich, *Fed. Proc.* **45**, 1913 (1986); K. King, S. J. Benkovic, P. Modrich, *J. Biol. Chem.* **264**, 11807 (1989).
13. J. P. Priestle, *J. Appl. Crystallogr.* **21**, 572 (1988).
14. We thank R. Unger and J. Sussman for calculating the subunit decomposition of our structure. Supported by NIH grant GM25671 and grant DMB890026P for computer time at the Pittsburgh Supercomputing Center.

3 July 1990; accepted 16 August 1990