Perspective

How Eco RI Recognizes and Cuts DNA

HELEN M. BERMAN

THE CRYSTAL STRUCTURE OF THE RESTRICTION ENZYME endonuclease RI bound to its cognate DNA is the first determination of a structure of a complex between an enzyme and the specific piece of DNA that it recognizes and cleaves; as such it represents a milestone in modern molecular biology. It is, therefore, worth reviewing the background that led to this particular achievement.

The birth of molecular biology occurred in 1953 when Watson and Crick (*1*), using data obtained by Rosalind Franklin and Maurice Wilkins (*2*), proposed the now well-known structure of DNA. That structure, of course, allowed the correct speculation that genetic information, stored in the base sequence, could be passed along via the very specific pairing of nucleic acid bases. What is not as widely known and appreciated is that this structure was determined with the use of fiber diffraction and required a great deal of chemical intuition; it did not allow for a direct unbiased view at high resolution. That achievement did not occur until 1973, when the first crystallographic determination of the double helix at atomic resolution was made (*3*). It should, perhaps, be noted that John Rosenberg, the principal investigator of the Eco RI complex (*4*), was a participant in that study.

Since that time there have been approximately 50 single-crystal determinations of DNA fragments, both alone and complexed with small molecule drugs. Most of these fragments are very small—dinucleoside phosphates. Among the largest is a dodecamer which, fortuitously, contains the Eco RI recognition site. In addition to this single-crystal work, the structures of numerous defined sequence polymers of DNA have been determined by means of fiber methods (*5*).

The most important conclusion that can be made from a review of this body of work is that DNA is polymorphous. It can have the right-handed B form (*6*). It can form short squat A helices (*7*), and it can even form left-handed Z helices (*8*). Furthermore, when complexed with simple small molecules it can do very peculiar and unexpected things, such as forming kinks and bends and even non-Watson-Crick base pairs (*9*). Simple changes in temperature or ionic strength or solvent are enough to change the structure. Unfortunately, at this time one cannot predict how a particular base sequence or environmental factor will affect the structure. However, with the recent availability of pure homogeneous material, we can expect to see an explosion of new structural information that will allow us to understand the principles underlying the detailed structure of this key molecule.

Protein crystallography has a much longer history and can now be said to have come of age. There are well over 300 structures reported since the first such determination, that of myoglobin (*10*), more than 25 years ago. These include such proteins as small and elegant as crambin (*11*) and as large and complex as the photosynthetic reaction center (*12*). Methods have been developed to allow for rapid data collection of high molecular weight samples, and supercomputers can be used for data processing. Just as important has been the increasing availability of large quantities of pure protein, due in part to the use of recombinant DNA technology to clone the genes and allow for their overexpression. How fitting then that the "star player" in this technological revolution be the subject of the crystallographic analysis reported here. Indeed, it is the use of restriction enzymes, of which Eco RI was the first to be discovered, that has allowed the crystallographer to become increasingly problem oriented and unrestrained by the shortage of highly purified samples.

One of the critical problems that has been the subject of crystallographic investigations over the last 10 years has been how particular proteins interact with DNA in specific and nonspecific ways. This knowledge is necessary if one is to understand replication, transcription, and translation. Nonspecific DNA-protein interactions occur, for example, in the nucleosome core particle, where DNA wraps around the histone octamer; the structures of both the core particle (*13*) and the histone octamer (*14*) have been determined. Very specific interactions occur between segments of DNA and proteins, such as repressors and activators. In these cases there have been several studies completed of the proteins themselves (*15*); the interactions between the DNA and proteins have been predicted from molecular modeling calculations based on the assumption that neither the DNA nor the protein undergoes any gross conformational change when they get together. In one case—that of the complex between the 434 repressor and its operator sequence (*16*)—this assumption does not have to be made, although at the resolution of 7 Å it is not really possible to define the detailed features of the complex. Thus the structure determination of Eco RI–DNA at 3 Å resolution is a truly welcome event.

What do we learn from this structure? Not unexpectedly, the binding of the protein to the DNA maintains twofold symmetry. This type of symmetry was first seen in the binding of a nucleic acid base to the drug actinomycin (*17*) and was correctly predicted as being a key feature of all DNA-protein interactions. We also see that the DNA changes from the canonical B form to a kinked helix when it is complexed with the enzyme. Kinks have been postulated and even seen in other structures, but the exact form that they take in any particular situation is not yet predictable. This is the first example of an enzyme-induced kink.

The results of this determination also imply that the protein must change from its native form, although the exact nature of that change must await determination of the structure of the protein alone. The protein in the complex exhibits the now familiar motifs, with separate but interrelated regions responsible for recognition and cutting of the DNA. Interestingly, one of the motifs has the pattern first identified by Rossmann (*18*) as the nucleotide binding fold. We also learn that, unlike the repressor proteins that have a helix-turn-helix pattern involved in the recognition of their cognate DNA's, in this structure a single unitary helix is responsible for recognition. The reason for this, according to Rosenberg, is that the repressors are smaller proteins and hence need more "structure" to anchor the recognition helix. Every part of Eco RI is called into play to perform the anchoring function. For dessert, we have the wonderful description of the 12 hydrogen bonds involved in the sequence specificity and the knowledge that, while these types of

The author is a senior member of the Institute for Cancer Research, Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111.

hydrogen bonds had been predicted (*19*) as the key to such specificity, the exact hydrogen bonds that come into play here could not have been predicted without knowledge of the exact geometry of the complexed DNA. This should serve as ample warning of the dangers of modeling without all the players being well known and characterized.

Crystals, in so far as they are ordered, are essentially static. Hence the argument has been made that the molecules contained within them may not resemble the same molecules in solution or in the biologically active state, and that one might have a difficult time understanding enzymology by examining crystal structures. The results of this determination, combined with the biochemistry done by the coauthors and others, considerably increase our understanding of how this enzyme actually works. The crystal itself was deliberately constructed to be inactive by removing Mg^{2+} from the crystallization solution. It is thus assumed that the active site is not fully assembled. There is a solvent channel in the area of the active site where it is highly probable that Mg^{2+} enters. Moreover, it is possible to make the crystals active and still maintain crystallographic diffraction by diffusing in Mg^{2+}. Furthermore, the analysis of the hydrogen bond specificity can account for the splicing activities under nonstandard conditions dubbed Eco RI* specificity. This article thus presents a very convincing case that crystallography can indeed explain enzymology.

This article also brings up some methodological issues that are worth bearing in mind more for what is unsaid than for what is actually discussed. The success of this determination was dependent not only on large quantities of pure material but on choosing just the right length of DNA to form well-ordered crystals. This has been found to be the case in other determinations as well. The problem is that we do not know ahead of time what length is needed. There were great difficulties in handling the crystals. They were small, weakly diffracting, and very radiation-sensitive. To date, crystallographers have found crystals to be idiosyncratic in their behavior, with some people and their samples being lucky and other not. This is changing, and there is now a heavy emphasis on quantitating and demystifying the process. Indeed, there has already been one workshop on crystal growth (*20*); and the next American Crystallographic Association Meeting (Austin, Texas, March 1987) has several sessions where this subject will be discussed. The actual structure determination was greatly facilitated by the use of ISIR methodology, which has become increasingly important as a means of deriving structures. It is also worth noting that the use of sophisticated molecular graphics as a tool for both the analysis and display of the structure is almost taken for granted, although it is interesting that at one point in the analysis the investigators use the older tried and true method of visualizing the electron density maps on plexiglass sheets. Finally, it should be noted that the collaborative efforts of crystallographers, biochemists, and molecular biologists from different laboratories hastened the technical analysis and helped to clarify the functional issues in an expeditious manner.

What next? First of all, this structure will be subject to continued refinement. Although no one can predict the results, it would not be at all surprising to find even greater deviations from the canonical B-DNA. The complex is ripe for suitably cautious modeling of the interactions with different lengths of DNA. In hand are crystals of the protein alone as well as of the protein with the cleavage products and, of course, this structure now becomes a strong candidate for site-directed mutagenesis. When all this is done, we will be in a position to understand more fully how this particular restriction enzyme does its work. Moreover, the analysis of another restriction enzyme will have the benefits of the lessons learned with this one. The ultimate and achievable aim is that in time we will indeed be able to really understand the rules that govern the way DNA is recognized and cleaved by other molecules.

REFERENCES

1. J. D. Watson and F. H. C. Crick, *Nature (London)* **171**, 737 (1953).
2. R. E. Franklin and R. G. Gosling, *Acta Cryst.* **6**, 673 (1953).
3. J. M. Rosenberg, N. C. Seeman, J. J. P. Kim, F. L Suddath, H. B. Nicholas, A. Rich, *Nature (London)* **243**, 150 (1973).
4. J. A. McClarin, C. A. Frederick, B.-C. Wang, P. Greene, H. W. Boyer, J. Grable, J. M. Rosenberg, *Science* **234**, 1526 (1986).
5. S. Arnott, *Prog. Biophys. Mol. Biol.* **21**, 267 (1970); A. G. W. Leslie, S. Arnott, R. Chandrasekaran, R. L. Ratliff, *J. Mol. Biol.* **143**, 49 (1980).
6. R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura, R. E. Dickerson, *Nature (London)* **287**, 755 (1980).
7. O. Kennard, *Pure Appl. Chem.* **56**, 989 (1984).
8. A. H.-J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel, A. Rich, *Nature (London)* **282**, 680 (1979).
9. G. Ughetto, A. H.-J. Wang, G. J. Quigley, G. van der Marel, J. H. van Boom, A. Rich, *Nucleic Acids Res.* **13**, 2305 (1985).
10. J. C. Kendrew, "Myoglobin and the Structure of Proteins" (Nobel Lecture, 11 December 1962), in *Nobel Lectures in English* (Elsevier, New York, 1963).
11. W. A. Hendrickson and M. M. Teeter, *Nature (London)* **290**, 107 (1981).
12. J. Deisenhofer, O. Epp, K. Miki, R. Huber, H. Michel, *ibid.* **318**, 618 (1985).
13. T. J. Richmond, J. T. Finch, B. Rushton, D. Rhodes, A. Klug, *ibid.* **311**, 532 (1984).
14. R. W. Burlingame, W. E. Love, B.-C. Wang, R. Hamlin, N. H. Xuong, E. N. Moudrianankis, *Science* **228**, 546 (1985).
15. W. F. Anderson, D. H. Ohlendorf, Y. Takeda, B. W. Matthews, *Nature (London)* **290**, 754 (1981); C. O. Pabo and M. Lewis, *ibid.* **298**, 443 (1982); D. B. McKay and T. A. Steitz, *ibid.* **290**, 744 (1981); R. W. Schevitz, Z. Otwinowski, A. Joachimiak, C. L. Lawson, P. B. Sigler, *ibid.* **317**, 782 (1985).
16. J. E. Anderson, M. Ptashne, S. C. Harrison, *ibid.* **316**, 596 (1985).
17. H. M. Sobell, S. C. Jain, D. D. Sakore, C. E. Nordman, *Nature New Biol. (London)* **231**, 200 (1971).
18. M. G. Rossmann, A. Liljas, C-I. Branden, L. J. Banaszak, "Evolutionary and structural relationships among dehydrogenases," in *The Enzymes*, P. O. Boyer, Ed. (Academic Press, New York, ed. 3, 1975), vol. 11, p. 61.
19. N. C. Seeman, J. M. Rosenberg, A. Rich, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 804 (1976).
20. *Protein Crystal Growth* (First International Conference on Protein Crystal Growth, Stanford University, Stanford, CA, 14 to 16 August 1985) (North-Holland, Amsterdam, 1986).

8 November 1986; accepted 14 November 1986