Acta Cryst. (1995). D**51**, 938–945

Preliminary Characterization of *Eco*RI*–DNA Co-crystals: Incomplete Factorial Design of Oligonucleotide Sequences

BY PATRICIA A. WILKOSZ, K. CHANDRASEKHAR AND JOHN M. ROSENBERG†

Departments of Biological Sciences and Crystallography, University of Pittsburgh, Pittsburgh PA 15260, *USA, and the W. M. Keck Center for Advanced Training in Computational Biology, University of Pittsburgh, Carnegie Mellon University, and the Pittsburgh Supercomputing Center, Pittsburgh PA* 15260, *USA*

(*Received* 29 *July* 1993; *accepted* 15 *March* 1995)

Abstract

A full understanding of the sequence specificity of *Eco*RI endonuclease requires structural information on complexes where the DNA contains one 'incorrect' base pair; historically, these sites are referred to as *Eco*RI* sites. They are inherently asymmetric, unlike the canonical *Eco*RI site, GAATTC, which possesses a twofold axis of rotational symmetry. All previously determined DNA–*Eco*RI complexes incorporated this symmetry axis into the space group, requiring the design of 'new' oligonucleotides to produce an asymmetric unit appropriate to an *Eco*RI* complex. The incomplete factorial approach of Carter & Carter [Carter & Carter (1979). *J. Biol. Chem.* **254**, 12219–12223.] was used to design the DNA sequence. Factors included the location and type of *Eco*RI* substitution and the length and AT richness of the sequences on both sides of the RI site. Co-crystals were obtained with several sequences, including one with TCGTG**GACTTC**GTG. Diffraction data were collected from one crystal of this complex to 3.2 Å resolution; the unit-cell parameters are $a = b = 123.8$ and $c = 148.9$ Å and the space group is $P3_221$. Unit-cell and space-group information was also obtained for the *Eco*RI* sites AAATTC, GGATTC and GAGTTC. These experiments demonstrated the need for a rapid, economical method that would distinguish DNA–protein co-crystals from crystals of protein only. This can be readily achieved with a single small crystal and a staining method based on methylene blue and methyl violet, which stain DNA and protein, respectively.

Introduction

*Eco*RI endonuclease normally recognizes its target sequence GAATTC with very high fidelity; double-strand cleavage of DNA at incorrect sites is virtually undetectable under physiological conditions. However, this specificity is relaxed under certain non-physiological buffer conditions, known historically as *Eco*RI* conditions, wherein sites containing one incorrect base pair (referred to as *Eco*RI* sites) are cleaved at detectable rates.

In principle, there are 18 *Eco*RI* sites, however they are equivalent in pairs because the *Eco*RI recognition site is an 'inverted repeat' (and because of the base-pairing rules and polarity of DNA). That is, the following two *Eco*RI* duplexes are identical,

```
~AAATTC~          ~GAATTT~
 ||||||     =      ||||||
~TTTAAG~          ~CTTAAA~
```

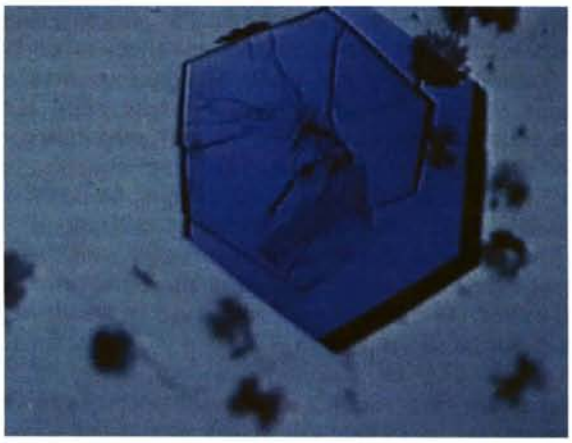
There are nine unique *Eco*RI* sites: AAATTC, TAATTC, CAATTC, GGATTC, GCATTC, GTATTC, GAGTTC, GACTTC and GATTTC. All of these sites are cleaved under '*Eco*RI* conditions', albeit at different rates. Note that the nine unique sites can be written such that the substitution (in the 'upper' strand) is within the first three base pairs

Solution studies have provided key data on the sequence-specific recognition of *Eco*RI endonuclease because they reveal the functional consequences of every possible base-pair substitution within the recognition sequence (Lesser, Kurpiewski & Jen-Jacobsen, 1990; Rosenberg, 1991). The biochemical studies are provocative, but difficult to interpret in the absence of structural studies on *Eco*RI* complexes. Here we report the preliminary characterization of co-crystals containing most of these complexes

These crystallizations shared a problem common to all DNA–protein co-crystallizations: The design of appropriate oligonucleotides. This is often difficult because a large number of possible sequences should be investigated while practical factors especially the cost of synthesis, limit the number of oligonucleotides that can be used. This is no different from the general problem of designing crystallization experiments where the number of conceivable conditions vastly exceeds the number of experiments that can be realistically performed. Carter & Carter developed an incomplete factorial approach to investigating crystallization conditions (Carter & Carter, 1979); here we report the application of the method to oligonucleotide design.

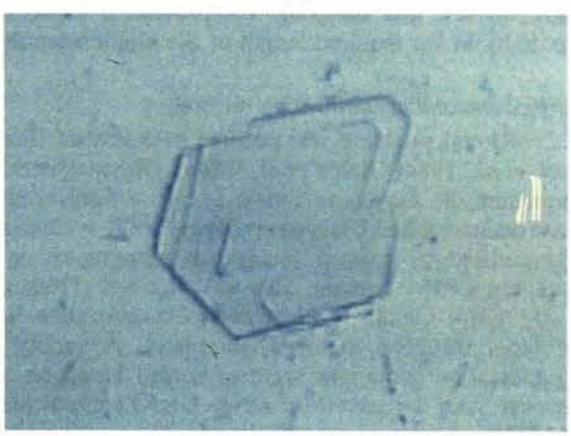
† To whom correspondence should be addressed at the Department of Biological Sciences, University of Pittsburgh, Pittsburgh PA 15260, USA.

ISSN 0907-4449



(*a*)

(*b*)



(*c*)

Fig. 1. Dye tests applied to *Eco*RI endonuclease crystals. The figure shows the results of staining *Eco*RI crystals with methylene blue and methyl violet, which stain DNA and protein, respectively. (*a*) Staining of a DNA–*Eco*RI endonuclease co-crystal with methylene blue. The crystals stained more intensely than the photograph suggests although not as strongly as with methyl violet. Salt crystals do not stain with either dye. (*b*) Staining of a similar co-crystal with methyl violet. (*c*) Staining of a crystal of the protein alone (without DNA) with methylene blue; these crystals did not stain. This demonstrates that methylene blue can be used to discriminate between protein crystals that do or do not contain DNA.

We also encountered the problem of distinguishing between protein crystals and DNA–protein co-crystals (not to mention distinguishing these from salt crystals). Electrophoretic methods of determining the composition of crystals are somewhat time consuming and require either a large crystal or several smaller ones. Large crystals are usually highly valuable in the initial stages of a crystallization experiment and there is considerable reluctance to sacrifice one or more for electrophoretic analysis.

As reviewed by McPherson (McPherson, 1982), dye tests were used to discriminate protein crystals from salt near the turn of the century (Hopkins & Pinkus, 1898; Sumner, 1919). Although this method hasn't been popular in more recent times, McPherson suggested that methyl violet and methylene blue could be used to discriminate protein and DNA crystals, respectively, from salt.

Here we also report a simple procedure for the use of these stains. These dye tests are relatively quick and easy and require one relatively small crystal each.

Methods

*Incomplete factorial design strategy for Eco*RI* complexes*

Canonical DNA–*Eco*RI endonuclease complexes possess a twofold axis of rotational symmetry that has been incorporated into the crystallographic lattices of the co-crystalline forms studied to date (Frederick *et al.*, 1984; Grable *et al.*, 1984; McClarin *et al.*, 1986; Kim, Grable, Love, Greene & Rosenberg, 1990; Kim *et al.*, 1994). However, *Eco*RI* substitutions are inherently asymmetric; a typical example is GACTTC. The A to C substitution destroys the inverted repeat (twofold symmetry) of the canonical sequence, GAATTC. Hence if an *Eco*RI* co-crystal were isomorphous to the canonical co-crystal then the *Eco*RI* co-crystal would necessarily be disordered. Thus, crystallization of *Eco*RI* complexes requires a larger asymmetric unit (at least a dimer) in order to visualize the structural changes produced by the base substitution.

Lesser *et al.* showed *via* ethylation interference that the entire DNA–protein interface becomes asymmetric in *Eco*RI* complexes (Lesser *et al.*, 1990). It is conceivable, therefore, that an appropriate co-crystal could be obtained simply by including an *Eco*RI* site in an oligonucleotide. However, it should be noted that the lattice contacts in the canonical co-crystal do not involve the regions of the complex where movement was suggested by the ethylation interference data, and we were not overly optimistic that this simple method would work.

Our approach to this oligonucleotide design problem was based on patterns we have noted in the packing of all *Eco*RI co-crystals obtained to date. The molecular arrangement along the direction of the average DNA

Table 1. *Factors used in the incomplete factorial approach to oligonucleotide sequence design*

Factor	Values
Position of *Eco*RI* substitution†	First, second, or third base pair
Type of *Eco*RI* substitution	A, T, C if the substitution is at the 1st base pair
	G, C, T if the substitution is at the 2nd to 3rd base pair
Length of 'left' flanking sequence	3,4 or 5 base pairs
AT/GC content of 'left' sequence	AT rich; AT poor
Lenth of 'right' flanking sequence	3,4 or 5 base pairs
AT/GC content of 'right' sequence	AT rich; AT poor

† Note that the *Eco*RI site (GAATTC) is six base pairs long but because of its internal symmetry, there are only nine unique *Eco*RI* substitutions and they can all be written such that it occurs on the 'left': AAATTC, TAATTC, CAATTC, GGATTC, GCATTC, GTATTC, GAGTTC, GACTTC, GATTTC.

helix axis is sensitive to the length of the oligonucleotide employed because DNA–DNA contacts are important components of the overall packing. However, protein–protein contacts control the packing in the plane perpendicular to the DNA helix axis; here the packing is very similar to that all of the previous cases (Grable, 1990). This suggested that placing the *Eco*RI* site off center in the oligonucleotide would produce the required asymmetry.

It would be prohibitive to examine all possible combinations of the nine unique *Eco*RI* sites described above. We have also obtained (canonical) co-crystals with 12-mers, 13-mers and 15-mers with evidence that other lengths could yield co-crystals as well (Grable, 1990). The unit cell and space group appear to be determined primarily by the length of the oligonucleotide while the quality of the diffraction pattern is strongly influenced by both the overall length of the oligonucleotide and by the base sequence within the regions flanking the *Eco*RI site.

We reduced the number of oligonucleotides to 21 duplexes by the incomplete factorial approach of Carter & Carter (1979), as implemented in a locally written computer program (Rosenberg, unpublished). The following factors were varied (see Table 1), the *Eco*RI* substitution, the length of each region flanking the *Eco*RI* site, and the A/T richness of each flanking region.

Two factors were required to vary the *Eco*RI* substitution economically; one was the position of the change within the *Eco*RI site. As noted, the *Eco*RI* sequences can all be written such that the substitution is within the first three base pairs. Hence, three values are required to describe the position of the substitution. The second factor was the nature of the base-pair change itself as described in Table 1; this also required three values. It would have required approximately three times as many oligonucleotides to treat all nine *Eco*RI* sites as a single factor and this would have been prohibitively expensive.

For both the 'left' and 'right' flanking sequences, we varied the length from three to five base pairs. (Left and right are defined with the *Eco*RI* substitution on the left.) We used two types of flanking sequence, one AT rich and the other poor in AT content, based on our previous experience with *Eco*RI co-crystallizations.

The Carter & Carter procedure begins with the abstract representation consisting of all combinations of all values of the factors; in this case it would be a six-dimensional space with $3 \times 3 \times 3 \times 2 \times 3 \times 2 = 324$ cells. 21 cells are picked at random. Some of the choices are then modified to insure that every pairwise combination of factors and values was tested at least once (see Carter & Carter, 1979, for a more complete discussion of this 'balancing' step).

We included 'dangling T's' (unpaired 5′ thymidines) in the oligonucleotides because they had demonstrated that they could assume at least two different positions in co-crystals. The DNA duplexes stack end-to-end in both previously determined *Eco*RI co-crystals and the formation of those continuous DNA rods appeared to be a prerequisite to the formation of good co-crystals. The dangling T's were stacked upon each other in the 13-mer complex described above producing an effective length of 14 stacked bases or base pairs (Kim, 1990). The unpaired thymidines were 'swung out' of the stack in the 15-mer complex, which also had 14 stacked bases as the effective asymmetric unit of the DNA rods (Grable, 1990). (Here, the unpaired thymines contact the protein.) Thus we felt that dangling T's provide a degree of flexibility in the required length of the oligonucleotide.

Purification and crystallization methods

*Eco*RI endonuclease was purified as described (Frederick *et al.*, 1984; Grable *et al.*, 1984). Oligonucleotides containing the sequences shown in Fig. 4 were synthesized on an Applied Biosystems Model 380B synthesizer and purified by reverse phase chromatography on a Hamilton PRP-1 column, also as described (Frederick *et al.*, 1984; Grable *et al.*, 1984). Crystallization was by vapor diffusion in Cryschem plates. A variety of crystallization conditions were generated based on experience with the growth of DNA–*Eco*RI endonuclease and DNA–topoisomerase co-crystals.

Dye test for protein–DNA co-crystals

Methylene blue and methyl violet were obtained from Fisher Scientific and Coleman and Bell, respectively. They were dissolved in distilled water at concentrations of 35 and 25 mg ml^{-1}, respectively; this gives a very dark solution. A small volume (0.3 µl) of dye solution was added to a crystallization well that contained approximately 10 µl of mother liquor and crystals. Positive results could sometimes be noted within 15 to 30 min,

however the wells were monitored over a 2 d period and additional 0.3 µl aliquots of dye added every few hours until the result was clear.

The concentration of dye in the mother liquor at the end of the reaction is critical because the reaction is judged by the contrast between the crystal and the mother liquor. The mother liquor should appear lightly colored when viewed through a dissecting microscope. Under these conditions a positively staining crystal will appear noticeably darker than the surrounding mother liquor. This contrast is difficult to detect if the mother liquor is too darkly stained and it is better to err on the side of introducing too little dye than too much. (A positively staining crystal can remove a considerable amount of dye from the mother liquor).

Crystal size is also critical. We have found that co-crystals need to be at least 100–150 mm long to stain reliably with methylene blue while somewhat smaller crystals suffice for methyl violet. This may also depend on crystal composition; smaller crystals containing only DNA may work for methylene blue.

One correction for having added too much dye is to remove some mother liquor, however the depth of mother liquor must still be greater than that of the crystals for the test to be valid. If the concentration of dye is appropriate a salt crystal will appear subtly less colored than the mother liquor. The best approach is to adjust the volume of the aliquots of dye solutions until clear discriminations are obtained with crystals of known composition.

Indeed, it cannot be emphasized strongly enough that this method must be practised on crystals of known composition in order to determine the correct amounts of dye to use and to 'calibrate the experimenter' so that she or he can recognize the staining differences. Once this is done, these dye tests are quite reliable.

Results and discussion

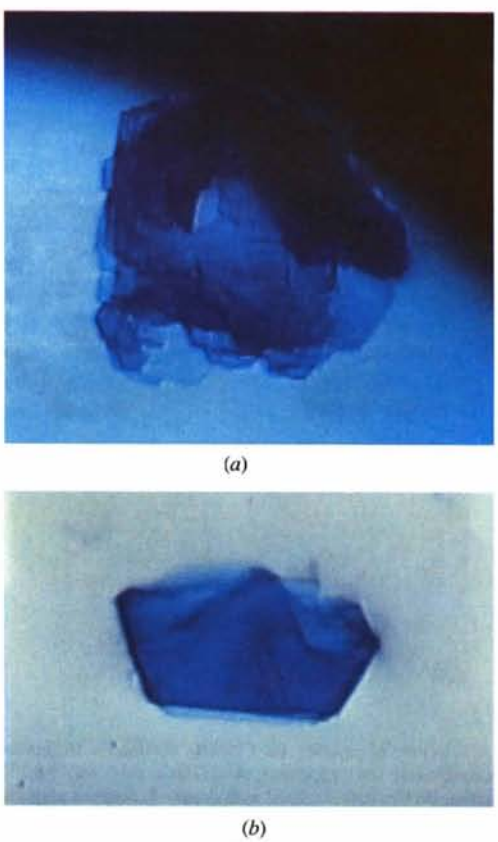
Dye tests

We tested the indicator dyes on two crystalline forms of *Eco*RI endonuclease of known composition: One is a DNA–protein co-crystal and the other a crystal of the protein alone. The co-crystal contains the tridecamer, TCGC**GAATTC**GCG which includes the canonical *Eco*RI recognition site (in bold); it is the crystal whose structure has been reported (Kim *et al.*, 1990; Rosenberg, 1991; Kim *et al.*, 1994). Fig. 1 shows that methyl violet (the protein stain) deeply stains the co-crystal, as expected. Methyl violet also stains crystals of the protein alone (not shown). Methylene blue (the DNA stain) also stained co-crystals, see Fig. 1, but it did not stain the crystals of the protein alone. Thus, the staining pattern with these dyes is sufficient to discriminate between DNA–protein co-crystals and crystals containing only protein. The staining with methyl violet is more pronounced perhaps because the crystals contain proportionately more protein than DNA (by a mass ratio of approximately 8:1).

*Eco*RI* *co-crystallizations*

The aim of these experiments was a search for the sequence and length of the 'flanking sequence' environment that is appropriate for the requirements of *Eco*RI* co-crystals; *i.e.* the asymmetry described above. An additional goal was to investigate the role of DNA sequences flanking the *Eco*RI site because it has been demonstrated that different flanking sequences cause the DNA cleavage rate to vary by at least an order of magnitude (Thomas & Davis, 1975). This effect is magnified at (some) *Eco*RI* sites.

It was expected that crystals would be obtained from solutions containing oligonucleotides but that many of them would be crystals of the protein alone. However, most of the crystals that we obtained stained positively with methylene blue indicating that they contained DNA (see Fig. 2 for representative examples). They all stained positively with methyl violet hence they are co-crystals.



(*a*)

(*b*)

Fig. 2. Dye tests applied to *Eco*RI duplex 17 crystals. (*a*) Crystals stained with methyl violet. These are badly aggregated. (*b*) A crystal stained with methylene blue. The crystal stains positively with methylene blue demonstrating the presence of DNA.

The figure suggests that the quality of most of these initial crystallization results appears disappointing; however, we did not use the best crystals for this test because it is destructive: Fig. 3 shows a more representative sampling of the crystals obtained in the first round of this experiment.

Entries in Fig. 4 with solid double boxes indicate oligonucleotides from which crystals diffract to at least 3.5 Å resolution. At the time of writing, we are still optimizing crystallization conditions, however our previous experience with *Eco*RI endonuclease crystallizations suggests that further improvement in resolution should be possible.

Some of the crystals indicated in Fig. 4 and shown in Figs. 2 and 3 are indistinguishable in habit from those obtained from the protein alone. We, therefore, expected that they would not contain DNA. Surprisingly, some of them stained positively with methylene blue (and methyl violet) while others did not. Preliminary precession photographs of crystals grown under conditions that gave methylene blue-positive results showed that although the unit cell appears unchanged from that of the protein alone, the distribution of intensities is notably different (see below). We, therefore, believe that they are DNA–protein co-crystals that have a unit cell that is identical (or very similar) to that of the protein alone. Preliminary data sets were collected on either a Siemens X-100 or Rigaku R-AXIS area detector for duplexes 4, 10, 15, 17 and 19. Unit-cell and space-group data for these forms are shown in Table 2.

The crystals obtained with duplex 17 were notably better at the outset although we suspect the optimal conditions for this duplex happened to be close to one of the first conditions employed. Initial crystals were obtained at pH 6.3 using 100 m*M* MES buffer and 7.5% MPD at 277 K. Since the dye tests demonstrated that they were co-crystals, we focused our efforts on them. An adjustment of the crystallization conditions enabled us to obtain single crystals over 0.8 mm long; the current conditions are equilibration *versus* pH 5.5, 6% MPD at 283 K.

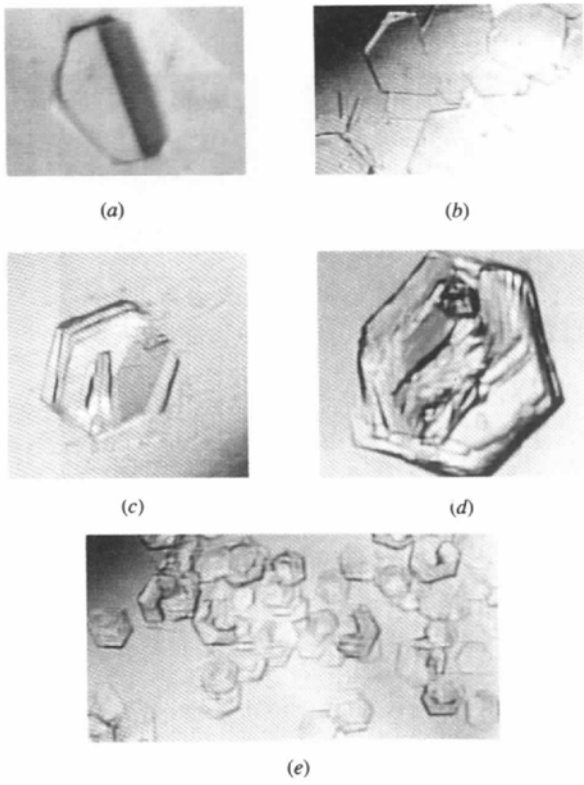


Fig. 3. A selection of crystals. (*a*) Crystals obtained with duplex 17, it is approximately 0.4 mm across. A similar crystal was used for data collection. (*b*) Crystals obtained with duplex 2, they are approximately 0.3 mm across; their habit is similar to that obtained with the protein alone. (*c*) Crystals obtained with duplex 15, they are approximately 0.45 mm across, however they are not single crystals. (*d*) Crystals obtained with duplex 16, they are approximately 0.5 mm across. (*e*) Crystals obtained with duplex 18, they are approximately 0.15 mm across.

Duplex 17 *data collection, unit cell and space group*

A diffraction data set to 3.2 Å resolution was measured from one duplex 17 crystal on a Siemens X-100 area detector. The unit-cell parameters are: $a = b = 123.8$, $c = 148.9$ Å, $\alpha = \beta = 90$, $\gamma = 120°$; and the space group

Fig. 4. Oligonucleotides employed in the crystallization experiments. An oligonucleotide in a dotted box produced co-crystals with *Eco*RI endonuclease; an oligonucleotide in a solid double box yielded co-crystals that gave useful X-ray diffraction data.

Table 2. *Eco*RI *endonuclease lattice parameters*

Description†	DNA	Crystal system space group	Cell parameters (Å,)	Current resolution (Å)
DNA *Eco*RI* complex	Duplex 4 (AAATTC)‡	Monoclinic *C*2	$a = 208.9$, $b = 126.5$, $c = 50.3$, $\beta = 99.9$	3.6
DNA *Eco*RI* complex	Duplex 10 (GGATTC)	Monoclinic *C*2	$a = 208.1$, $b = 126.4$, $c = 49.8$, $\beta = 99.4$	3.5
DNA *Eco*RI* complex	Duplex 15 (GACTTC)	Monoclinic *C*2	$a = 209.3$, $b = 126.5$, $c = 50.3$, $\beta = 99.9$	3.2
DNA–*Eco*RI* complex	Duplex 17 (GFACTTC)	Trigonal *P*3,21	$a = b = 123.8$ $c = 148.9$	3.2
DNA–*Eco*RI* complex	Duplex 19 (GAGTTC)	Monoclinic *C*2	$a = 208.1$, $b = 126.4$, $c = 49.8$, $\beta = 99.4$	3.4
'Original' DNA *Eco*RI complex	13-mer (GAATTC)§	Trigonal *P*321	$a = b = 118.4$ $c = 49.8$	2.5
'Apo' *Eco*RI endonclease	None	Monoclinc *C*2	$a = 208.6$, $b = 126.9$, $c = 49.7$, $\beta = 98.4$	3.0
Methyl DNA–*Eco*RI complex	Methylated 13-mer	Monoclinic *C*2	$a = 207.3$, $b = 127.3$, $c = 49.4$, $\beta = 99.6$	2.8
DNA *Eco*RI complex	15-mer (GAATTC)¶	Rhombohedral	$a = b = 127.8$ $c = 147.2$	3.5

† The first five entries in the table represent the new crystal forms described here; the remaining entries have been obtained previously in our laboratory and are being shown for comparison purposes.

‡ Duplex refers to the entry number in Fig. 4; the *Eco*RI* site is shown with the substitution emphasized.

§ The complete sequence of this oligonucleotide is TCGCGAATTCGCG.

¶ The complete sequence of this oligonucleotide is TCGTGGAATTCCACG.

is either $P3_121$ or $P3_221$. The data were processed and scaled with *XENGEN* (Howard *et al.*, 1987); the overall scaling R factors are 10.5% (on I) and 5.3% (on F). The only other space group consistent with the symmetry of reciprocal space is $P321$, which may be eliminated by the systematic absences along c^*. Data statistics are summarized in Table 3. Fig. 5 shows pseudo-precession photographs generated from this data. Additional data are being collected from several crystals and we anticipate a final resolution in the neighborhood of 3.2 Å.

We have obtained a molecular-replacement solution in space group $P3_221$ using the 'invariant' core of the protein as a search model (Wilkosz *et al.*, unpublished). The core included only those portions of the protein that are well removed from the DNA–protein interface; protuberances and other surface features likely to vary from crystal to crystal were also deleted from the model. Omitted regions, including DNA, were apparent in the electron-density maps and we are currently completing the refinement of this structure.

Duplex 15 *data collection and comparison with protein in the absence of DNA*

A similar data set was collected on the Siemens X-100 for duplex 15, which is also a GACTTC site. The unit-cell parameters are $a = 209.3$, $b = 126.5$, $c = 50.3$ Å, $\alpha = \beta = 90$, $\gamma = 99.9°$, and the space group is $C2$. These parameters are very similar to those of the 'apo protein' *i.e.* that obtained in the absence of any oligonucleotide (see Table 2). The habits of these two crystal forms are also indistinguishable. However, the crystals obtained with duplex 15 stain positively with methylene blue while those of the apo protein do not.

Table 3. *Duplex* 17 (GACTTC) *preliminary data-collection statistics*

Average resolution (Å)	No. possible	No. observed	R factor (on I)
8.1	3949	2969	6.0
5.1	3777	2084	10.1
4.3	3765	1375	10.2
3.8	3724	656	10.2
3.5	3715	170	9.9

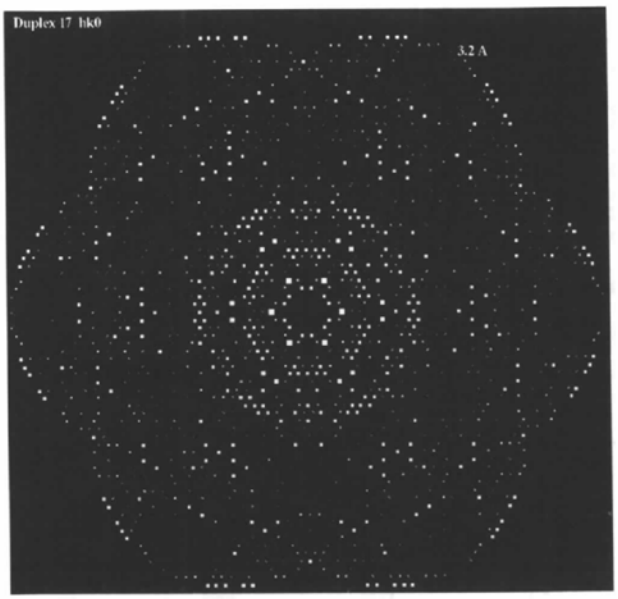
Fig. 6 shows a comparison of the $hk0$ zones of the diffraction patterns obtained from duplex 15 and apo protein crystals. Although the distribution of intensities is similar, many differences can be noted. We conclude that the packing and overall spatial arrangement of the protein is similar in the two cases with the diffraction differences coming from the presence of DNA in the former, coupled with local adjustments of the protein in the vicinity of the DNA.

We have recently determined the structure of the apo protein and the regions of the protein that would bind to the DNA do not participate in lattice contacts (Chandrasekhar *et al.*, in preparation). Indeed, the DNA-binding site is occupied by bulk solvent in the apo protein lattice and the 'hole' is large enough to accommodate a double helical fragment the size of duplex 15.

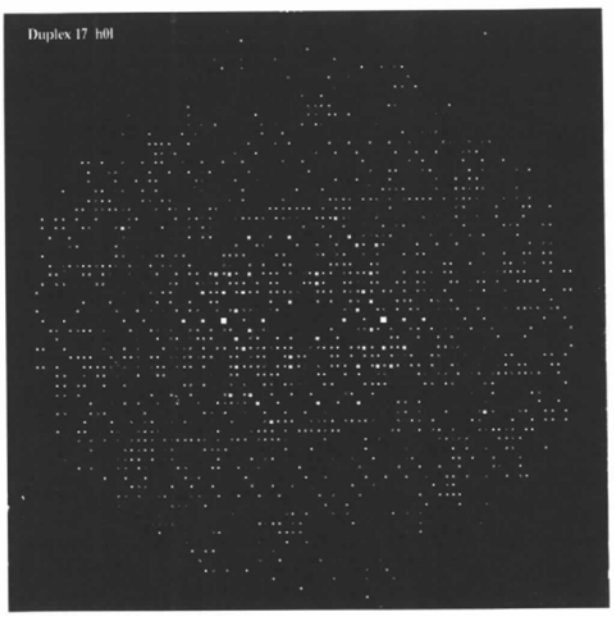
The packing arrangement of the apo protein in the monoclinic lattice is strikingly similar to that of the protein in the 'original' $P321$ co-crystal. This is possible, despite the apparent difference in space groups, because the asymmetric unit of the apo protein contains three subunits of the endonuclease that are related by a non-crystallographic threefold axis of rotational symmetry. Furthermore, only the ends of the DNA participate

in lattice contacts in the original co-crystal; similarly, the regions of the protein that bind the DNA do not participate in lattice contacts in that form either.

These fortuitous circumstances allow a comparison of protein and protein–DNA complexes within the same basic lattice environment. This should facilitate comparisons with biochemical experiments. It also points out the utility of the dye tests presented here. We almost ignored many DNA–protein complexes because they appeared to be identical to the apo protein; indeed, without the dye tests we probably would not have pursued these forms.

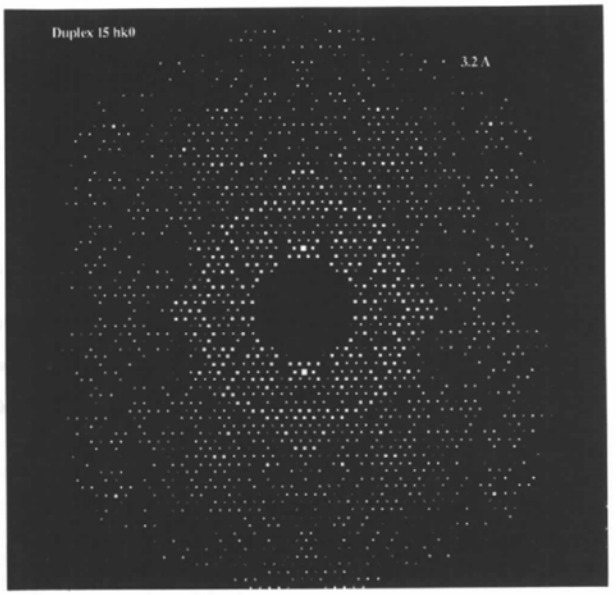


(*a*)



(*b*)

Fig. 5. Pseudo precession photographs of the co-crystals obtained with duplex 17: The figures were generated from the *XENGEN* data set summarized in Table 2 using the program *PRECESS* on a Silicon Graphics 4D25 workstation. The limit of resolution in both cases is 3.2 Å. (*a*) The *hk*0 zone. (*b*) The *h*0*l* zone (note that the *c* axis is horizontal).



(*a*)

(*b*)

Fig. 6. Pseudo precession photographs of the co-crystals obtained with duplex 15 and of the apo protein. These figures were generated similarly to those of Fig. 5. (*a*) The *hk*0 zone of the Duplex 15 co-crystals, which are in space group *C*2. (*b*) The same zone from *XENGEN* data obtained from crystals of the 'apo protein' which is the protein crystallized in the absence of any DNA. The differences are consistent with the presence of DNA in the former, which stains positively with methylene blue.

This work was supported by a grant from the National Institute of General Medicine (GM25671) and by the W. M. Keck Center for Advanced Training in Computational Biology at the University of Pittsburgh, Carnegie-Mellon University and the Pittsburgh Supercomputer Center.

References

CARTER, C. W. JR & CARTER, C. W. (1979). *J. Biol. Chem.* **254**, 12219–12223.

FREDERICK, C. A., GRABLE, J., MELIA, M., SAMUDZI, C., JEN-JACOBSON, L., WANG, B. C., GREENE, P., BOYER, H. W. & ROSENBERG, J. M. (1984). *Nature (London)*, **309**(5966), 327–331.

GRABLE, J. C. (1990). *Preliminary Structure of a R32 EcoRI Endonuclease–DNA Complex*. Univ. of Pittsburgh, USA.

GRABLE, J., FREDERICK, C. A., SAMUDZI, C., JEN-JACOBSON, L., LESSER, D., GREENE, P., BOYER, H. W., ITAKURA, K. & ROSENBERG, J. M. (1984). *J. Biomol. Struct. Dynam.* **1**(5), 1149–1160.

HOPKINS, F. G. & PINKUS, S. N. (1898). *J. Physiol.* **23**, 130–136.

HOWARD, A. J., GILLILAND, G. L., FINZEL, B. C., POULOS, T. L., OHLENDORF, D. H. & SALEMME, R. (1987). *J. Appl. Cryst.* **20**, 383–387.

KIM, Y. (1990). *Crystal Structure Analysis of EcoRI Endonuclease–Substrate DNA Complex and EcoRI Endonuclease–Product Complex*. Univ. of Pittsburgh, USA.

KIM, Y., CHOI, J., GRABLE, J. C., CHANDRASEKHAR, K., WILKOSZ, P. A., GREENE, P., HAGER, P. & ROSENBERG, J. (1995). *Structure*. Submitted.

KIM, Y. C., GRABLE, J. C., LOVE, R., GREENE, P. J. & ROSENBERG, J. M. (1990). *Science*, **249**, 1307–1309.

LESSER, D. R., KURPIEWSKI, M. R. & JEN-JACOBSON, L. (1990). *Science*, **250**, 776–786.

MCCLARIN, J. A., FREDERICK, C. A., WANG, B. C., GREENE, P., BOYER, H. W., GRABLE, J. & ROSENBERG, J. M. (1986). *Science*, **234**, 1526–1541.

MCPHERSON, A. (1982). *Preparation and Analysis of Protein Crystals*. New York: John Wiley.

ROSENBERG, J. M. (1991). *Curr. Op. Struct. Biol.* **1**, 104–113.

SUMNER, J. B. (1919). *J. Biol. Chem.* **37** 137.

THOMAS, M. & DAVIS, R. W. (1975). *J. Mol. Biol.* **91**, 315–328.