The *in Vitro* Assembly of the *Eco*KI Type I DNA Restriction/Modification Enzyme and Its *in Vivo* Implications†

David T. F. Dryden,*,‡ Laurie P. Cooper,‡ Peter H. Thorpe,‡ and Olwyn Byron§

Institute of Cell & Molecular Biology, The University of Edinburgh, Darwin Building, The King's Buildings, Mayfield Road, Edinburgh, EH9 3JR, U.K., and National Centre for Molecular Hydrodynamics, Adrian Building, University of Leicester, University Road, Leicester, LE1 7RH, U.K.

Received August 5, 1996; Revised Manuscript Received November 21, 1996⊗

ABSTRACT: Type I DNA restriction/modification enzymes protect the bacterial cell from viral infection by cleaving foreign DNA which lacks N6-adenine methylation within a target sequence and maintaining the methylation of the targets on the host chromosome. It has been noted that the genes specifying type I systems can be transferred to a new host lacking the appropriate, protective methylation without any adverse effect. The modification phenotype apparently appears before the restriction phenotype, but no evidence for transcriptional or translational control of the genes and the resultant phenotypes has been found. Type I enzymes contain three types of subunit, S for sequence recognition, M for DNA modification (methylation), and R for DNA restriction(cleavage), and can function solely as a M_2S_1 methylase or as a $R_2M_2S_1$ bifunctional methylase/nuclease. We show that the methylase is not stable at the concentrations expected to exist *in vivo*, dissociating into free M subunit and M_1S_1, whereas the complete nuclease is a stable structure. The M_1S_1 form can bind the R subunit as effectively as the M_2S_1 methylase but possesses no activity; therefore, upon establishment of the system in a new host, we propose that most of the R subunit will initially be trapped in an inactive complex until the methylase has been able to modify and protect the host chromosome. We believe that the *in vitro* assembly pathway will reflect the *in vivo* situation, thus allowing the assembly process to at least partially explain the observations that the modification phenotype appears before the restriction phenotype upon establishment of a type I system in a new host cell.

Protein−protein interactions between subunits of an oligomer or transient contacts between different proteins are vital to many cellular processes such as the control of enzymatic activity, signalling pathways, DNA synthesis, and transcription (Klotz, 1975; Phizicky & Fields, 1995). The strength of these interactions ranges from very weak with a dissociation constant, K_d,[1] in the millimolar range to very strong with a K_d in the picomolar range. The strength of the interactions required to produce a particular reaction may also reflect the concentration of the proteins in a cell or cell compartment with abundant proteins, such as glycolytic enzymes, interacting more weakly, and rare proteins, such as transcription factors, having to interact more strongly to remain bound to their targets once they have been located in the crowded intracellular space (Srivastava & Bernhard, 1987; Goodsell, 1991; Reddy *et al*., 1995).

Bacterial type I DNA restriction/modification (R/M) systems are large oligomeric proteins which, by recognizing the pattern of adenine methylation in specific DNA sequences, can switch activity between either cleavage of unmethylated, foreign DNA, thereby restricting its propagation, or methylation of hemimethylated DNA produced by DNA replication to fully modify it before the next cell division. The oligomeric nature of type I R/M systems plays an important role in switching between these two different activities via a complex series of conformational changes in response to the methylation state of the DNA target sequence (Bickle, 1993; Bickle & Kruger, 1993; King & Murray, 1994).

The 20 known type I systems have been found primarily in *Escherichia coli* and *Salmonella enterica*, but it seems likely that they are widespread as examples have recently been found or postulated in a variety of Gram positive and Gram negative bacteria and in an archaebacterium (Dybvig *et al*., 1994; Fleischmann *et al*., 1995; Stein *et al*., 1995; Valinluck *et al*., 1995; Xu *et al*., 1995; Bult *et al*., 1996). Type I R/M systems are the products of the *hsd* (*h*ost *s*pecificity for *D*NA) *R*, *M*, and *S* genes which code for the restriction (R), modification (M), and sequence specificity (S) subunits, respectively. They have been grouped into four

† This work was supported by grants from the Medical Research Council, Biotechnology and Biological Sciences Research Council, Engineering and Physical Sciences Research Council, and The Royal Society. P.H.T. received a Ph.D. Studentship from the Medical Research Council, and D.T.F.D. was the recipient of a University Research Fellowship from The Royal Society.

* Correspondence should be addressed to this author. Phone: +44-(0)131-650-5378. FAX: +44(0)131-650-8650. Email: David.Dryden@ed.ac.uk.

‡ The University of Edinburgh.

§ University of Leicester.

⊗ Abstract published in *Advance ACS Abstracts,* January 15, 1997.

[1] Abbreviations: ATP, adenosine 5′-triphosphate; bp, base pair(s); DEAE, diethylaminoethyl; EDTA, ethylenediaminetetraacetic acid; HPLC, high-performance liquid chromatography; hsd, host specificity of DNA; IPTG, isopropyl β-D-thiogalactopyranoside; K_d, dissociation constant; kb, kilobase(s); kDa, kilodalton(s); MES, 2-(*N*-morpholino)-ethanesulfonic acid; mtase, DNA methyltransferase; mod+/−, modification proficiency or deficiency; PAGE, polyacrylamide gel electrophoresis; PMSF, phenylmethanesulfonyl fluoride; res+/−, restriction proficiency or deficiency; R/M, restriction/modification; rpm, revolutions per minute; SAM, *S*-adenosylmethionine; SDS, sodium dodecyl sulfate; TBE, 0.045 M Tris−borate, 1 mM EDTA, pH 8 buffer; Tris, 2-amino-2-(hydroxymethyl)-1,3-propanediol.

S0006-2960(96)01943-5 CCC: $14.00

families, IA−ID, based on *in vivo* complementation between systems containing defective subunits, antibody cross-reactivity, DNA hybridization, and in some cases amino acid sequence similarity (Barcus & Murray, 1995; Barcus *et al*., 1995). The *in vivo* complementation suggests that at least within a family, the structures of subunit interfaces are well conserved and that subunits can be easily exchanged during assembly of the complete enzyme. This level of structural conservation is maintained between species that are evolutionary distant.

The biochemical activities displayed by the proteins formed by the *hsd R*, *M*, and *S* genes have been fairly well characterized [reviewed in Modrich (1979), Yuan (1981), and Bickle (1993)], and simple structural models based on subunit stoichiometries (Burckhardt *et al*., 1981), sequence homology with the small type II mtases of known structure (Dryden *et al*., 1995), and domain swapping experiments between type I systems (Gann *et al*., 1987; Cowan *et al*., 1989; Kneale, 1994) have been described. One S subunit is found in each complete protein and is essential for recognizing the DNA target. The target for the S subunit of *Eco*KI from *E. coli* K12 is AAC $(N)_6$ GTGC. Two M subunits, responsible for DNA modification of the adenine bases at the underlined positions in the target, bind to one S subunit to form an active mtase even in the absence of the R subunit (Lautenberger & Linn, 1972; Suri *et al*., 1984; Nagaraja *et al*., 1985; Dryden *et al*., 1993; Taylor *et al*., 1993). The mtases of the *Eco*KI (Dryden *et al*., 1993) and *Eco*R124I (Taylor *et al*., 1993) systems methylate hemimethylated target sequences approximately 100-fold faster than unmodified target sequences, a preference which is also found in the complete bifunctional type I system. The binding of two R subunits to the type I mtase completes the type I R/M system (Eskin & Linn, 1972; Meselson *et al*., 1972; Suri *et al*., 1984; Nagaraja *et al*., 1985) except for the *Eco*R124I system where only one R subunit appears to be present (Janscak *et al*., 1996). On binding to an unmodified target, a complex series of events ensues which results in DNA cleavage after extensive DNA translocation driven by ATP hydrolysis [reviewed in Modrich (1979), Yuan (1981), and Bickle (1993)]. The R subunits are essential for the cleavage process and contain all of the amino acid motifs found in the DEAD box family of DNA and RNA helicases and putative helicases (Gorbalenya & Koonin, 1991; Webb *et al*., 1996). Despite their complexity, type I systems such as *Eco*KI are very efficient in restricting the propagation of phage with typical cutbacks in phage propagation of between 10^3 and 10^5, with 10^8 being possible in some circumstances (Webb *et al*., 1996).

It has long been known that the genes for a type I system can be transferred successfully to an unmodified host despite the expectation that the expression of restriction would be suicidal for the unmodified chromosome (Arber & Dussioux, 1962; Boyer, 1964; Colson *et al*., 1965; Fuller-Pace *et al*., 1985; Suri & Bickle, 1985; Loenen *et al*., 1987; Prakash *et al*., 1991; Prakash-Cheng & Ryu, 1993; Prakash-Cheng *et al*., 1993; Redaschi & Bickle, 1996). Since these transfers can occur, there must exist some way of preventing the restriction phenotype from appearing before or at the same time as the modification phenotype. No evidence for transcriptional or translational control has ever been found, and it has been suggested that some form of post-translational control of the expression of restriction exists to account for the observations that restriction appears a considerable time after the appearance of modification.

In this paper, we will describe the *in vitro* assembly of the *Eco*KI type I R/M system from its constituent subunits and how this assembly pathway may account for the observed regulation of restriction and modification phenotypes of type I systems *in vivo*.

MATERIALS AND METHODS

The expression vector pJF118EH, in which transcription of the cloned *hsd* genes is from the IPTG-inducible tac promoter, was used to amplify the levels of both the mtase and the M subunit. pJFMS, which includes the *hsdM* and *hsdS* genes on a 7.1 kb *Sma*I−*Eco*RI fragment, has been described (Dryden *et al*., 1993). *hsdM* was overexpressed from the plasmid pJFM created by ligating a 2.1 kb *Sma*I−*Hin*dIII fragment from pJFMS into pJF118EH. The R subunit was obtained by using the multicopy plasmid pJK2 containing *hsdR* and its natural promoter, p^{res}, in vector pAT153 (Kelleher *et al*., 1991). Our attempts to overexpress *hsdR* from an inducible promoter have so far been unsuccessful. pRH3, a 6.16 kb derivative of pBR322 containing the 1.8 kb *Hin*dIII fragment from the *hsd* region of *E. coli* K12, was used as a substrate for the *Eco*KI endonuclease (Sain & Murray, 1980). Large-scale preparations of pRH3 containing either unmodified *Eco*KI sites or fully modified *Eco*KI sites were made using *E. coli* strains NM679 and C600, respectively (Clewell & Helinski, 1969). pRH3 contains *Eco*KI sites at 3458 and 5831 bp orientated from 5′ to 3′ and 3′ to 5′, respectively, an *Eco*RI site at 0 bp, and a *Pvu*II site at 3867 bp. DNA oligonucleotide (OSWEL) duplexes containing either 25 or 45 bp and one central *Eco*KI site either unmodified or hemimethylated (Powell *et al*., 1993) were used for methylation assays.

Mtase, M_2S_1, was purified as described (Dryden *et al*., 1993) but substituting the S300 Sephacryl gel filtration column with a Superdex 200 16/60 column (Pharmacia). M_1S_1 was also obtained as a byproduct of mtase preparation and was obtained in an apparently homogeneous form by careful selection of chromatographic fractions in the final DEAE-Sepharose and Superdex gel filtration steps; however, when assayed for mtase activity, it was found to have approximately 5% of the activity of the same molar amount of mtase.

M subunit was prepared from 5 L of *E. coli* NM522/pJFM grown at 37 °C in L broth containing 0.1 mg/mL ampicillin. IPTG was added to 1 mM when the culture reached an optical density of 0.5 at 650 nm. Twelve grams of cell paste was harvested 4 h after induction of transcription. Coomassie blue staining of 10% polyacrylamide−SDS gels clearly showed the M subunit and these gels were used to detect the M subunit through purification. The following procedures were at 4 °C. The paste was resuspended in 55 mL of 20 mM Tris, 20 mM MES, 10 mM $MgCl_2$, 7 mM β-mercaptoethanol, 1 mM EDTA, 10^{-5} M PMSF, and 10^{-5} M benzamidine, pH 8, and sonicated for 15 min with intermittent cooling on ice. Cell debris was removed by centrifugation at 27000*g* for 3 h. The supernatant was applied at 48 mL/h to a DEAE-Sepharose fast flow 33 cm × 1.6 cm diameter column (Pharmacia) and eluted with a 500 mL 0−0.5 M NaCl gradient at 24 mL/h in the above buffer. The protein in fractions containing M was precipi-

tated by adding ammonium sulfate to 80% saturation. The precipitate (~420 mg total of which 200 mg was lost accidentally) was resuspended in 12 mL of the above buffer containing 2 M NaCl and applied to a 90 cm × 2.6 cm diameter Sephacryl S300 gel filtration column at 24 mL/h. The 100 mg of protein recovered was again precipitated with ammonium sulfate, and the precipitate was dialyzed against buffer containing no NaCl before reapplication to the DEAE ion exchange column. A 500 mL gradient of NaCl from 0 to 0.4 M at 24 mL/h did not give a satisfactory improvement in purity, so the M subunit containing fractions were pooled, precipitated, and dialyzed as above before application to a BioRad Econo Q 5 mL anion exchange column at 48 mL/h. A 120 mL 0−0.5 M NaCl gradient gave sufficient purification to allow the recovery of ~50 mg of M subunit that was judged to be approximately 95% pure as determined by Coomassie blue staining of SDS gels. The M subunit can be stored for over 3 years as a precipitate in 80% ammonium sulfate at 4 °C.

The R subunit was prepared from 5 L of NM679/pJK2 grown in L broth containing 0.1 mg/mL ampicillin for 6 h at 37 °C. Thirty grams of cell paste was resuspended in 100 mL of 20 mM Tris-HCl, 10 mM $MgCl_2$, 7 mM β mercaptoethanol, 10% glycerol, 10^{-5} M PMSF, and 10^{-5} M benzamidine, pH 7.5, buffer and sonicated on ice with intermittent cooling for 20 min. Cell debris was removed by centrifugation for 1 h at 27000*g*. Solid NaCl was added to the supernatant to 0.4 M followed by the addition of 10% neutralized poly(ethylenimine) to 0.4% v/v. This was stirred for 15 min and the precipitated material removed by 30 min of centrifugation at 27000*g*. Solid ammonium sulfate was added to the supernatant to 70% saturation and stirred for 1 h, and the protein precipitate was recovered by centrifugation for 30 min at 27000*g*. The precipitate was resuspended in 50 mL of buffer and dialyzed against the same buffer for 3 h before being applied to the DEAE-Sepharose column. Bound protein was eluted with a 500 mL 0−0.5 M NaCl gradient. Fractions containing R in 0.2−0.25 M NaCl were dialyzed against buffer to remove NaCl and applied to a 17 cm × 1.6 cm diameter heparin−agarose column (Sigma). A 500 mL 0−1.0 M NaCl gradient eluted the R subunit at ~0.2 M NaCl. Fractions containing the R subunit were concentrated to 1 mL using a 10 kDa cutoff centrifugal concentrator (Flowgen) and applied to Superose 12 and Superose 6 FPLC gel filtration columns (Pharmacia) linked in series. Fractions containing >95% R subunit were obtained from these columns, concentrated, and stored at −20 °C after the addition of an equal volume of glycerol. A typical preparation yielded 2 mg.

*Eco*KI nuclease was prepared using a multicopy plasmid (a kind gift from A. Chen and N. Murray) expressing the *R*, *M*, and *S* genes in *E. coli* strain NM679. Six liters of culture grown in L broth supplemented with 0.1 mg/mL ampicillin at 37 °C yielded about 3 mg of homogeneous nuclease using the natural promoters for the *Eco*KI system. The plasmid construction and nuclease purification will be described in detail elsewhere, but briefly, the purification was similar to that of the R subunit with the substitution of the Superdex 200 16/60 column for the two FPLC Superose columns.

Concentrations of proteins were estimated using 280 nm molar extinction coefficients calculated from the tryptophan and tyrosine content of the proteins of 58 850, 84 289, 143 140, and 114 233 M cm^{-1} for the M, M_1S_1, M_2S_1, and R proteins, respectively (Sober, 1970). The Bradford assay for protein concentration was also used (BioRad). Nondenaturing and denaturing polyacrylamide gel electrophoresis was performed as described (Sambrook *et al.*, 1989). The denaturing gels for the resolution of high molecular weight complexes formed by glutaraldehyde cross-linking contained 0.8% agarose and 3.5% polyacrylamide. The agarose was dissolved in hot gel buffer and then the acrylamide added immediately prior to casting of the gel in a prewarmed gel apparatus. These gels were run without a stacking gel either in a Tris−glycine buffer (Sambrook *et al.*, 1989), with marker proteins of up to 205 kDa, or in 20 mM sodium phosphate, 2% SDS buffer, pH 7 (Sigma Technical Bulletin MWS-877X), with cross-linked phosphorylase *b* and cross-linked bovine serum albumin markers (Sigma). Glutaraldehyde cross-linking was performed at room temperature for 2 min by adding a 25% aqueous solution of glutaraldehyde to obtain a final concentration of 1% in 100 μL of a 0.2 mg/mL protein solution in the same buffer as used to prepare the M subunit. The reaction was terminated by the addition of 2.5 μL of 2 M $NaBH_4$ freshly prepared in 0.1 M NaOH, per 100 μL of sample for 20 min. Samples were then mixed with an equal volume of SDS−PAGE loading buffer and boiled for 5 min before electrophoresis. Protein was visualized by Coomassie blue staining (Sambrook *et al.*, 1989), silver staining (BioRad), or Western blotting using *Eco*KI antibodies (Murray *et al.*, 1982).

Mtase activity was assayed as described (Dryden *et al.*, 1993) with 0.5 μM protein, 10 μM DNA oligonucleotide duplex, and 50 μM SAM containing 1% ^{3}H-SAM (Amersham); 2 μL samples of a 25 μL reaction volume were counted. Nuclease activity was typically assayed at 25 °C in a 100 μL volume containing 10 nM *Eco*KI, 5 nM unmodified pRH3, 0.1 mM SAM, and 50 μg/mL bovine serum albumin. The buffer was 33 mM Tris−acetate,10 mM magnesium acetate, 66 mM potassium acetate, and 0.5 mM dithiothreitol, pH 7.9. The reactions were started by the addition of 40 mM stock solution of ATP to a final concentration of 2 mM; 10 μL samples were removed at set times, heated to 68 °C for 10 min, and mixed with 7 μL of 15% Ficoll in TBE buffer. The samples were loaded onto a 0.6% agarose TBE gel and electrophoresed in the TBE buffer. pRH3 plasmid was linearized if necessary with *Eco*RI or *Pvu*II endonuclease (New England Biolabs). These endonucleases were subsequently inactivated by heating to 68 °C for 10 min prior to the addition of *Eco*KI, serum albumin, SAM, and ATP.

HPLC gel filtration used a Rainin Dynamax 4.6 × 250 mm Hydropore-5-SEC column and a guard column; 20 or 50 μL samples in 20 mM Tris, 20 mM MES, 0.2 M NaCl, 10 mM $MgCl_2$, 7 mM β mercaptoethanol, and 0.1 mM EDTA, pH 6.5, were injected onto the column. The samples were equilibrated at 4 °C for 20 min prior to injection. The flow rate was 0.5 mL/min, and detection was most sensitive and stable at 254 nm. The column was calibrated with a series of globular proteins of known molecular weight, giving a linear calibration curve of log (molecular weight) as a function of elution time. Most of our samples, after buffer exchange into the column buffer using PD10 Sephadex G50 columns (Pharmacia), contained trace amounts of a small molecule which we believe to be glycerol. This gave rise to a "solvent peak" after approximately 6.5 min, the elution time of which served as an internal standard to correct for

slight run to run variation in protein elution times. The absorbance at elution times expected for the assembled complexes and the unassembled components was measured from the elution profiles. In the continuous variation titrations for the assembly of M with M_1S_1, the absorbance was converted to mtase concentration using a standard curve constructed by the injection of samples of known volume and concentration onto the column. The continuous variation titration method (Job, 1928; Agmus, 1961) involved the mixing of different amounts of M and M_1S_1 such that the total protein concentration was the same in all samples. In this method, the definition of the dissociation constant $K_d = [A][B]^n/[AB_n]$ for the equilibrium binding of n molecules of B by A is rewritten as $(C(1-\chi) - [AB_n])(C\chi - n[AB_n])^n = K_d[AB_n]$ where C is the total concentration of protein and χ the mole fraction of B in the sample. This has an analytical solution for $n = 1$, the situation encountered for the binding of M and M_1S_1; hence, the K_d could be determined by solving a quadratic equation for the amount of complex formed as a function of χ. We did not use this method for the binding of R to M, M_1S_1, or M_2S_1 but used instead a straightforward titration of a constant amount of protein with increasing amounts of R subunit. Data were fitted using the Grafit software package (Erithacus).

Dynamic laser light scattering was performed on a Protein Solutions DynaPro-801TC instrument at 20 °C using the same buffer as used in the HPLC experiments.

Sedimentation equilibrium was performed at 20 °C in a Beckman Optima XL-A analytical ultracentrifuge equipped with scanning absorbance optics. The buffer used was 20 mM sodium phosphate, 0.1 M NaCl, pH 8. Rotor speeds of 8000, 9000, or 10 000 rpm were used, and the solute distribution of mtase at equilibrium was recorded at wavelengths of between 265 and 278 nm depending on the loading concentration (0.05−5 μM). The solution column height was 1.5 mm, and equilibrium at each rotor speed was ascertained by overlaying scans acquired 3 h apart. A true optical base line, free of macromolecular solute, was obtained by overspeeding the rotor to 40 000 rpm. The apparent whole cell molecular mass was obtained by fitting the data with either the IDEAL1 model in the Beckman XL-A analysis software or the MSTAR program (Creeth & Harding, 1982; Harding *et al*., 1992).

RESULTS

Characterization of the M and R Subunits, M_1S_1, the Mtase, and the Nuclease. The *Eco*KI mtase and nuclease could be purified to apparent homogeneity by the methods described (Figure 1). Assuming that the amount of Coomassie blue staining of the protein subunits was proportional to their molecular weight, densitometry of these preparations indicated a 2:1 ratio of M subunit to S subunit for the mtase and 2:2:1 for R, M, and S subunits for the nuclease. The use of a higher resolution gel filtration medium to remove M_1S_1 and M more effectively during the mtase preparation can account for the improvement over the 1.7:1 ratio reported previously for the mtase (Dryden *et al*., 1993). The 2:2:1 ratio for the nuclease preparation agrees with that found for nuclease prepared by a different procedure (Weiserova *et al*., 1993). The subunit ratio for the purified M_1S_1 was 1:1, again reflecting an improvement over our previous method. The improved purification of M_1S_1 gave a preparation with

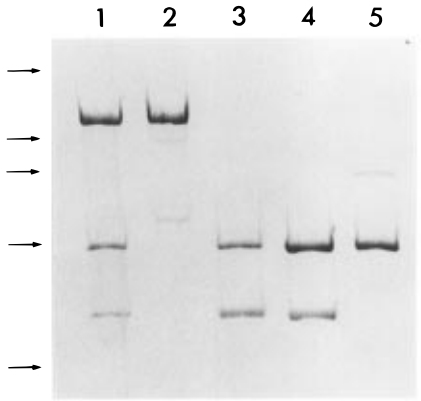

FIGURE 1: A 10% polyacrylamide, denaturing gel stained with Coomassie Blue showing the protein preparations used in this study. Lanes 1−5 show the *in vivo*-assembled nuclease, the R subunit, M_1S_1, M_2S_1 mtase, and the M subunit, respectively. The arrows indicate molecular mass markers of 205, 116, 97.4, 66, and 45 kDa from the top of the gel to the bottom.

less than 5% of the mtase activity of an equivalent concentration of mtase (data not shown). Earlier preparations had nearly 30% of the mtase activity, and we believe that this indicates that M_1S_1 has no intrinsic mtase activity (Dryden *et al*., 1993). The subunit compositions of the mtase, M_1S_1, and the nuclease agreed with the determination of molecular weight by a variety of methods (Table 1). SDS−PAGE of glutaraldehyde-cross-linked complexes of M_2S_1, M_1S_1, and the nuclease showed the formation of one major species in each case, but for the M_2S_1 and M_1S_1 preparations, a small amount of intermolecular cross-linking to give higher order forms was also apparent (Figure 2). These higher order assemblies were not detectable with any of the other methods and hence were most probably artifacts caused by the relatively high protein and glutaraldehyde concentrations used. The laser light scattering experiment gave translational diffusion coefficients of $(445.5 \pm 100) \times 10^{-13}$ m^2 s^{-1} and $(405.7 \pm 29) \times 10^{-13}$ m^2 s^{-1} for M_1S_1 and M_2S_1, respectively. Gel filtration appeared to overestimate the molecular weights of M_2S_1 and M_1S_1 which may reflect a slightly elongated shape.

The purification of the M and R subunits did not result in homogeneous protein samples despite considerable effort (Figure 1). This may be a reflection of a tendency for M and R to transiently associate with any contaminating proteins via exposed hydrophobic surfaces which would normally be buried in the assembled mtase and nuclease. We estimated that the M and R subunit preparations were about 95% pure. A Western blot of the R subunit preparation using polyclonal antibodies showed only one signal for the intact R subunit, indicating that none of the numerous contaminating proteins visible by silver staining were products of degradation of the R subunit (data not shown). N-Terminal sequencing of the R subunit revealed two overlapping N-termini which were different from that proposed from the nucleotide sequence of the *hsdR* region. Approximately two-thirds of the preparation began with the sequence MMNKS, the remainder with MNKSN, indicating that the coding region for the R subunit starts 61 base pairs downstream of that originally predicted (Loenen *et al*., 1987) and encodes a 134 kDa polypeptide. The subunit was a monomer in solution as determined by gel filtration and its inability to form higher molecular weight species in the presence of glutaraldehyde

Table 1: Apparent Molecular Masses (kDa) of Complexes Used in This Study As Determined by a Variety of Methods

	sample						
method	M	M_1S_1	M_2S_1	R	R + M	R + M_1S_1	R + M_2S_1
FPLC gel filtration	96	157	200[b]	101			
HPLC gel filtration	110	166	220	203	377−410	425−460	402−435
glutaraldehyde cross-link[a]	∼50	∼94	∼143	∼145	∼480	∼470	400−460
light scattering	polydisperse	110	160				
analytical ultracentrifuge			166[b]				
theoretical[c]	59	110	169	134	193	244	437

[a] Molecular mass of major species. [b] From Dryden *et al*. (1993). [c] Assuming M and R subunits are monomers and that the complexes of R with M, M_1S_1, and M_2S_1 have stoichiometries of R_1M_1, $R_1M_1S_1$, and $R_2M_2S_1$, respectively.

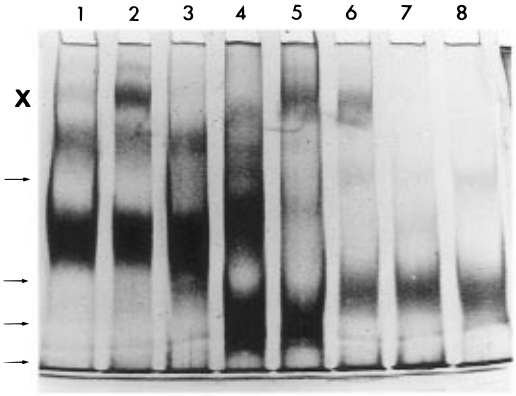

FIGURE 2: A 0.8% agarose−3.5% polyacrylamide denaturing gel of glutaraldehyde-cross-linked samples stained with silver after electrophoresis in a Tris−glycine buffer. The molecular mass markers indicated by arrows are 205, 116, 97.4, and 66 kDa from the top to the bottom. The dye front is just below the 66 kDa marker position. The X indicates the migration distance of the complexes between R and M_2S_1, M_1S_1, and M visible in lanes 2, 5, and 6. Lane 1: M_2S_1 mtase mostly as a cross-linked trimer but with some higher order complexes above 205 kDa due to intermolecular cross-linking also visible. Lane 2: M_2S_1 + R subunit showing the presence of a very high molecular mass species presumed to be the nuclease in addition to the species observed in lane 1. Lane 3: M_1S_1 + M subunit showing the formation of species identical to the mtase in lane 1. Lane 4: M_1S_1 mostly in a dimeric form at ∼100 kDa but with some intermolecular cross-linking also evident. Lane 5: M_1S_1 + R subunit showing the formation of a very high molecular mass species similar in size to the nuclease visible in lane 2. Lane 6: R subunit + M subunit showing the presence of a very high molecular mass species in addition to either dimeric M subunit or monomeric R subunit. Lane 7: M subunit cross-linked to form a dimer with the predominant monomeric form of the subunit having migrated off the bottom of the gel. Lane 8: R subunit present almost exclusively as a monomer.

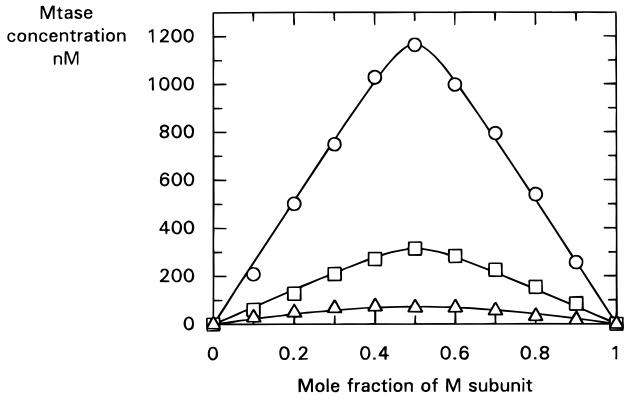

FIGURE 3: Apparent concentration of mtase formed by assembly of M_1S_1 with the M subunit as determined by the continuous variation titration method. The amount of material eluting from the gel filtration column at the time expected for the mtase was measured at total concentrations of M + M_1S_1 of 3 μM (○), 1 μM (□), and 0.3 μM (△). The data were fitted simultaneously to the equation given under Materials and Methods, with the K_d for the assembly of mtase from M_1S_1 and M constrained to be the same for each titration, but the total concentration allowed to vary independently. The best fits were found with a K_d of 15 ± 6 nM and total protein concentrations estimated to be 2.6, 0.75, and 0.21 μM for the 3, 1, and 0.3 μM titration, respectively.

concentrations sufficient to cross-link other oligomeric proteins such as the mtase (Table 1, Figure 2).

The M subunit was primarily a monomer in solution as determined by glutaraldehyde cross-linking despite its anomalous apparent molecular weight determined by gel filtration which was more suggestive of a dimer (Table 1). The anomalous behavior of the M subunit could indicate a rather elongated nonglobular structure. Gel filtration showed a broad high molecular weight shoulder of low intensity, which we interpreted as being a dimeric form of M, next to the principal monomeric form. Dimers and smaller amounts of higher order species were visible after glutaraldehyde cross-linking and SDS−PAGE; however, we estimated that approximately 90% of the M subunit was monomeric. Dynamic laser light scattering also indicated that the sample was polydisperse (data not shown). This polydispersity contrasts with the monodisperse, monomeric M subunit isolated from the *Eco*R124I type I system (G. Kneale, personal communication).

Mtase Assembly. Glutaraldehyde cross-linking between M and M_1S_1 clearly showed a complex of the same size as cross-linked mtase on SDS−PAGE, indicating the assembly of M with M_1S_1 to form M_2S_1 (Figure 2).

Due to the difficulty in obtaining quantitative information from the cross-linking experiments, we examined the assembly process in more detail using HPLC gel filtration and the method of continuous variation titration. In this titration, the mole fractions of M_1S_1 and M were varied, but the total concentration of protein was kept constant and higher than the K_d. The amount of material eluting from the gel filtration column at the apparent molecular weight of the M_2S_1 mtase was measured (Figure 3). The titrations clearly showed that the maximum amount of mtase was formed at a 1:1 ratio of M to M_1S_1. This also clearly indicated that the M subunit was predominantly monomeric. If the association of M and M_1S_1 was irreversible, then the titration would have shown a linear increase from the origin until $\chi_M = 0.5$, where the maximum amount of product was formed, and then a linear decrease to zero amount of product at $\chi_M = 1$. However, the titrations deviated from this form, and when all data sets (8 in total) were simultaneously fitted to the equation given under Materials and Methods, with the K_d constrained to be the same for all titrations, we obtained a K_d of 15 ± 6 nM for the equilibrium between M and M_1S_1 and the assembled

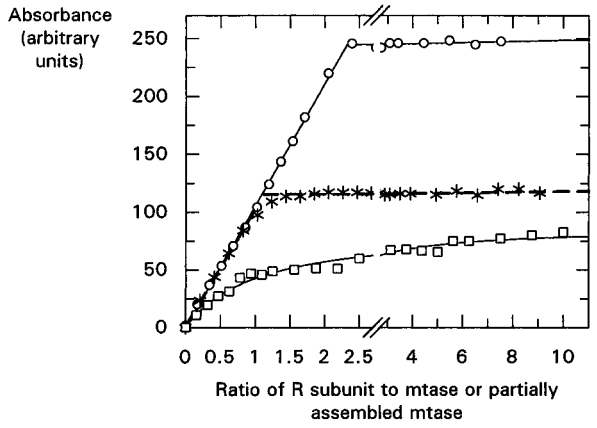

FIGURE 4: Titration of 750 nM M_2S_1 mtase (○), 750 nM M_1S_1 (∗), and 750 nM M subunit (□) with increasing amounts of the R subunit. The amount of material eluting with the high molecular masses given for the complexes of R with these proteins in Table 1 was determined by HPLC gel filtration. The titrations of M_2S_1- and M_1S_1 with R show a sharp end point at 2.3 R:M_2S_1 and 1.2 R:M_1S_1, indicating tight binding. The titration of the M subunit with R is indicative of weaker binding, and the curve drawn is for a K_d of 1 μM.

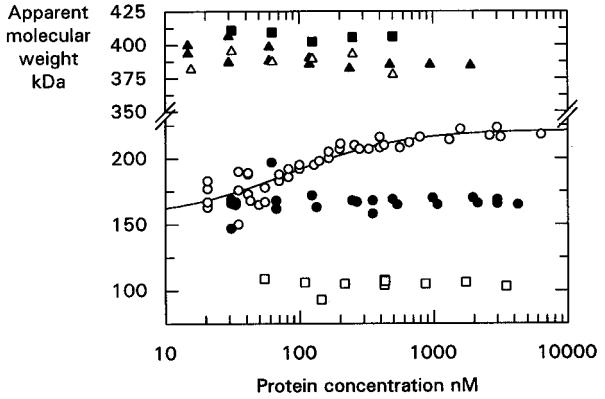

FIGURE 5: Apparent molecular masses of various *Eco*KI complexes determined from the elution time on an HPLC gel filtration column as a function of protein concentration. Data are shown for M_2S_1 (○), M_1S_1 (●), M subunit (□), *in vivo* assembled nuclease (▲), *in vitro* assembled nuclease (△), and the complex formed *in vitro* between the R subunit and M_1S_1 (■). The line fitted to the M_2S_1 data is for the dissociation of M_2S_1 with a K_d of approximately 80 nM.

M_2S_1 mtase. This K_d was within the range estimated by diluting the mtase (see later). Fitting of individual data sets gave values of K_d between 2.6 and 19 nM. The fitting procedure allowed the total concentration, [M] + [M_1S_1], to vary, and values of 2.6, 0.75, and 0.21 μM were obtained. These values were lower than the expected total concentrations of 3, 1, and 0.3 μM. The deviation between the expected concentration and that found by the data fitting increased as the concentration decreased. This may reflect the effects of zone spreading and separation of dissociated species by the gel filtration procedure. The assembly of M_2S_1 causes no major change in the conformation of the M subunit or of M_1S_1 as for both near- and far-UV circular dichroism, the spectrum of the mtase is equal to the sum of the spectra of the two components (unpublished observation with Prof. N. C. Price and Dr. S. Kelly, University of Stirling).

Assembly of Complexes with the R Subunit. SDS−PAGE of glutaraldehyde-cross-linked complexes of R and mtase (Figure 2) or of the *in vivo* assembled nuclease (data not shown) clearly showed the existence of a high molecular weight complex. High molecular weight complexes of R with M_1S_1 and M were also clear on SDS−PAGE of cross-linked samples (Figure 2).

Quantitative analyses of titrations of M, M_1S_1, and M_2S_1 with increasing amounts of the R subunit were performed with HPLC gel filtration (Figure 4). The binding between R and the mtase was so tight as to give stoichiometric binding with an end point of 2.3 R subunits per M_2S_1 mtase as determined by measuring either the amount of product eluting at a molecular mass of 400 kDa or the amount of excess R subunit eluting at 140 kDa (data not shown). No unbound, excess R subunit could be observed until the end point was reached. The slightly more than 2:1 R:M_2S_1 ratio may be attributed to an overestimation of the R subunit concentration as the preparation was not completely pure. The molecular mass of the complex varied between 402 and 435 kDa, close to the 440 kDa expected for an $R_2M_2S_1$ complex. No dissociation of the complex by dilution down to concentrations of approximately 20 nM could be observed by gel filtration (Figure 5). The *in vitro* assembled $R_2M_2S_1$ complex had the same elution volume as the nuclease assembled *in vivo* (Figure 5).

The association of R and M_1S_1 was also very tight with an end point being reached at a 1.2:1 ratio, and, once again, no free excess R subunit was observed until the end point was attained (Figure 4). The absorbance of the complex formed at the end point was 0.474 that of the complex of R and mtase. This value was close to the value of 0.534 expected from a consideration of the extinction coefficients of the subunits in the complexes. The molecular mass of this 1:1 complex varied between 425 and 460 kDa, similar to that found for the nuclease, so it probably has a subunit stoichiometry of $(R_1M_1S_1)_2$. Dilution of this complex to low concentrations did not alter its molecular mass, indicating that dissociation did not occur at concentrations above 20 nM (Figure 5).

The binding of R to M also produced a high molecular weight complex detectable by HPLC gel filtration; however, the binding was not stoichiometric, and the titration data were fitted to a one site ligand binding equation with a K_d of 1 μM for the association of R and M (Figure 4). Unbound R or M subunit could be observed throughout the titration in contrast to the other titrations. The theoretical end point of the titration at 86.5 units on the *y* axis was 0.347 of the absorbance at the end point of the R plus mtase titration. This was close to the value of 0.466 expected from the extinction coefficients of the R and M subunits and a 1:1 association. The molecular mass for the complex varied between 377 and 410 kDa, possibly consistent with a $(R_1M_1)_2$ stoichiometry.

Dissociation of the Mtase. One factor which can affect the stability of an oligomeric protein is its concentration in solution (Klotz, 1975). Dissociation at low concentrations will usually also abolish activity. The protein complexes which we have purified or assembled *in vitro* were examined by HPLC gel filtration as a function of concentration (Figure 5). Only the mtase showed any change in elution volume in the experimentally observable concentration range. The mtase appeared to dissociate into lower molecular weight species as its concentration was lowered, and although we cannot measure the apparent molecular weight of a complex

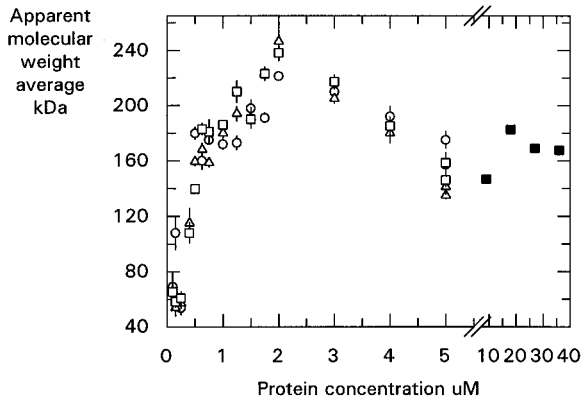

FIGURE 6: Apparent molecular mass of the mtase as a function of concentration measured by equilibrium sedimentation in an analytical ultracentrifuge. Data were measured at rotor speeds of 8000 rpm (○), 9000 rpm (△), and 10 000 rpm (□). Data shown as (■) are from Powell *et al*. (1993).

below ~25 nM, the K_d was in the 10−100 nM range. This was in reasonable agreement with the K_d found with the continuous variation titration method. As the concentration decreased, the peak of the elution profile shifted to lower apparent molecular weight with an increase in peak width due to a long trailing edge on the peak. It is most probable that the M_2S_1 mtase dissociates into one M subunit and one M_1S_1 dimer, the same species that can be isolated during the purification of the mtase. The S subunit of *Eco*KI is insoluble by itself (M. Winter, unpublished results) which may explain why the M_1S_1 dimer did not itself dissociate.

The dissociation of the mtase was confirmed by measurement of the average molecular weight by sedimentation equilibrium in the analytical ultracentrifuge as a function of mtase concentration which clearly showed that the apparent molecular weight average decreased rapidly for protein concentrations less than 500 nM (Figure 6). It was clear that the change in mass with concentration was a result of subunit association as opposed to polydispersity. Polydispersity would cause the mass at a given concentration to decrease with increasing rotor speed which it clearly did not. At high dilution, the smallest molecular weight species had a mass of ~60 kDa. The weight average mass increased to a maximum of ~220 kDa at 2 μM concentration and then decreased again at higher concentrations. The complexity of the mtase system makes these data hard to quantify explicitly; however, the behavior at low concentrations could suggest either that the S subunit is sparingly soluble and the M_2S_1 mtase can completely dissociate or that while M remains in solution M_1S_1 is unstable and oligomerizes and is pelleted out from the system. This explanation seems unlikely given the presence of species with mass in excess of 200 kDa at higher concentrations. The subsequent drop in apparent mass at increased concentration can be attributed either to nonideality, perhaps as a result of molecular asymmetry in the complex, or to the formation of higher oligomers which pellet from the system. Analysis of the data with the MSTAR program, which does not assume ideality, increased the apparent average molecular weight at each concentration but did not affect the overall shape of the graph.

Order of Subunit Assembly. Can the complexes of R with M, and R with M_1S_1 bind additional subunits to form the $R_2M_2S_1$ nuclease? This would be difficult to observe using

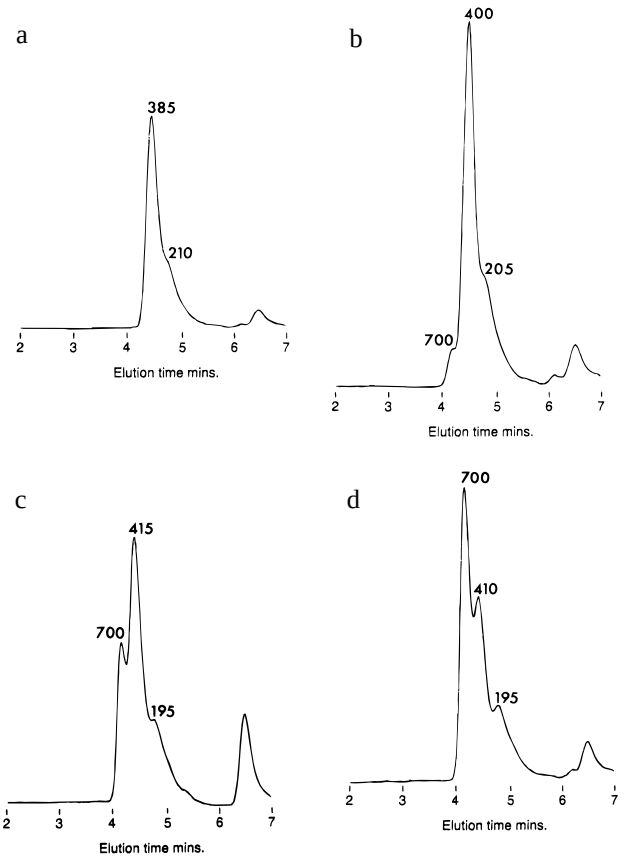

FIGURE 7: Elution profiles from the HPLC gel filtration column found by mixing various combinations of subunits. The molecular masses of the major peaks and shoulders are given in kilodaltons. The peak at approximately 6.5 min is a solvent peak used as an internal calibration to account for slight run to run variability in elution times. (a) 0.5 μM M_2S_1 + 1 μM R (same as 0.5 μM M+ 0.5 μM M_1S_1 and then, after 20 min incubation, +1 μM R); (b) 0.5 μM M + 0.5 μM M_1S_1 + 1 μM R mixed all at once; (c) 1 μM R + 0.5 μM M_1S_1 and then, after 20 min incubation, +0.5 μM M (same as 0.5 μM R + 0.5 μM M_1S_1 and then, after 20 min incubation, +0.5 μM R + 0.5 μM M); (d) 1 μM R + 0.5 μM M and then, after 20 min incubation, +0.5 μM M_1S_1.

the gel filtration or cross-linking methods as all of the complexes have approximately the same molecular weight; however, we have attempted to determine the most important steps in the assembly of the nuclease by mixing the subunits together in different orders and different proportions (Figure 7). The elution profiles of the subunit mixtures were different if the subunits were mixed in different orders. It was apparent that if preformed mtase or equal amounts of M subunit and M_1S_1 were not present on the addition of the R subunit, then a considerable amount of material with a molecular mass of ~700 kDa was formed in addition to complexes of ~400 kDa. If the R subunit was mixed with M subunit or M_1S_1 prior to the addition of further M_1S_1 or M subunit to give a 2:2:1 ratio of R:M:S subunits, then this aggregate was formed. Therefore, it appears that the most efficient method of making the nuclease is to make the mtase first and then add R subunits rather than proceed through complexes of R with M and M_1S_1. Whichever sequence of mixing R, M, and M_1S_1 was followed, a considerable amount of nuclease activity was detected, suggesting that even once R has bound to M or M_1S_1 it can dissociate to some limited extent not detectable by dilution of the complexes and HPLC gel filtration (data not shown). This possible dissociation of inactive complexes appears to be competing with an

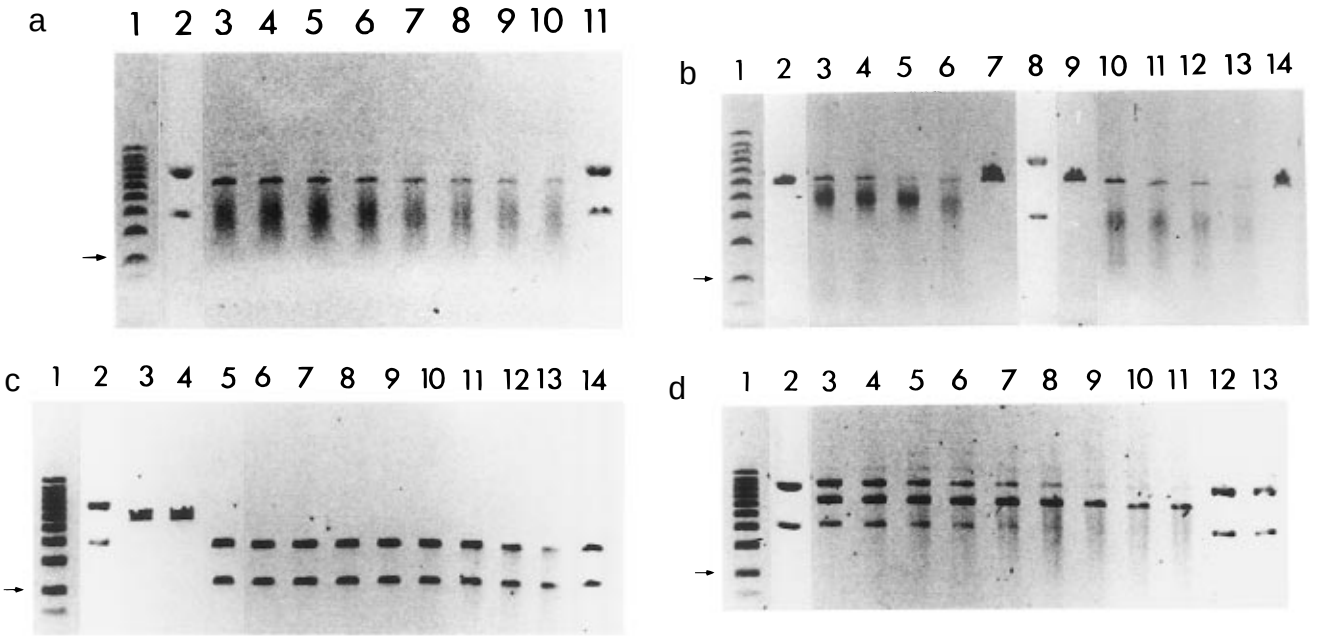

FIGURE 8: Time course of digestion of 5 nM pRH3, which contains two *Eco*KI target sites, as visualized by ethidium bromide staining of agarose gels. Lane 1 in all gels contains DNA size markers. The size marker indicated by the arrow is 2 kb in length, and all those above this marker increase by 1 kb for each band. The size marker visible below the 2 kb marker is 1.5 kb. The protein concentration was 10 nM in each panel. (a) Time course of digestion of circular pRH3, both supercoiled and nicked forms, by the *in vivo* assembled *Eco*KI nuclease. Lanes 2−10 are after 0, 0.5, 1, 1.5, 2, 5, 10, 20, and 30 min, respectively. Lane 11 shows the absence of cleavage after 30 min when no ATP is present. (b) Lanes 2−6 show the cleavage of pRH3 that has been linearized by *Eco*RI prior to incubation with *Eco*KI for times of 0, 0.5, 2, 5, and 30 min, respectively. Lane 7 is after 30 min exposure to *Eco*KI in the absence of ATP. Lane 8 shows the circular form of pRH3. Lanes 9−14 are the same as lanes 2−7 except that the pRH3 substrate was linearized with *Pvu*II prior to incubation with *Eco*KI. (c) Lane 2 shows circular pRH3; lanes 3 and 4 show pRH3 linearized by either *Eco*RI or *Pvu*II, respectively. Lanes 5−13 show the effect of *Eco*KI on pRH3 after it has been cleaved into two short pieces each now containing only one *Eco*KI site, by the action of both *Eco*RI and *Pvu*II at times of 0, 0.5, 1, 2, 5, 10, 20, 40, and 60 min. Lane 14 after 60 min in the absence of ATP. (d) This gel shows the effect of the protein assembled *in vitro* from M_1S_1 and the R subunit on the circular forms of pRH3. We believe the activity displayed is due to the slight contamination of M_1S_1 with the M_2S_1 mtase. Lanes 2−11 are after 0, 0.5, 1, 1.5, 2, 5, 10, 20, 40, and 60 min incubation with this protein. Lanes 12 and 13 are after 60 min incubation in the absence of SAM or ATP, respectively.

aggregation process facilitated by the addition of M or M_1S_1.

Nuclease Activity of Complexes Containing the R Subunit Assembled in Vitro Compared to the Nuclease Assembled in Vivo. The model (Studier & Bandyopadhay, 1988) of bidirectional DNA translocation past the enzyme with the *Eco*KI molecule remaining fixed at its initial target site and subsequent cleavage at a nonspecific sequence selected by the collision of two translocating type I nucleases approximately halfway between the initial target sites implies that *Eco*KI nuclease should produce two cuts in the circular pRH3 plasmid and one in the *Eco*RI or *Pvu*II linearized plasmid. The products of pRH3 cleavage should be two fragments of poorly defined, but not random, length. The cleavage reaction is unusual in that it does not show any turnover (Eskin & Linn, 1972). Cleavage of DNA containing two or more target sites is more efficient than cleavage of DNA containing only one site (Murray *et al*., 1973; Webb *et al*., 1996), fitting the proposed model which requires two enzymes to collide at the cleavage site (Studier & Bandyopadhay, 1988).

As noted above, the mixing of the correct ratios of R, M, and M_1S_1 in any order produced species of ~400 kDa. The most homogeneous preparations, as monitored by HPLC elution profiles, were formed when 2 equiv of R was added to 1 equiv of the preformed mtase or a 1:1 mixture of M and M_1S_1. We examined the nuclease activity of the *in vivo* assembled nuclease, of the *in vitro* assembled nuclease, and of complexes formed from R, M, and M_1S_1 in detail to confirm that the *in vitro* assembled nuclease was identical to that isolated from the *in vivo* expression system and that complexes of R with partially assembled M_1S_1 were inactive.

The nuclease assembled *in vivo* gave efficient cleavage of pRH3 when present at a concentration of one molecule per *Eco*KI target site. The cleavage was rapid and produced the fragments that would be predicted by the Studier and Bandyopadhay (1988) model whether one initially has a circular substrate or a linear one (Figure 8a,b). Cleavage of a circular substrate initially produced a linear, full-length DNA product, but further cleavage occurred to give a smear of products about 3 kb in length as expected. It was apparent that not all of the linearized DNA could be cleaved to completion as this form persisted for a considerable time. It is possible that the DNA translocation process prior to cleavage of circular DNA results in some full-length linear product which is topologically unsuitable for further cleavage. This was not the case if one used only the full-length linear substrate. This was rapidly digested to give a smear of products containing certain preferred lengths of cleavage products. The ~4.6 kb fragments produced by *Eco*KI action on *Eco*RI linearized plasmid appearred to be less diverse in size than the very faint ~1.6 kb fragments. We believe this is a reflection of the nonlinear rate of fragment migration through the gel depending on fragment length. Plasmid linearized by *Pvu*II nuclease was cleaved into approximately 2.4−3.8 kb fragments. As expected, no cleavage of the DNA was observed, even after 1 h, in the absence of SAM, ATP, or Mg^{2+}. No cleavage of plasmids prepared from an *Eco*KI modifying host was observed (data not shown). If pRH3

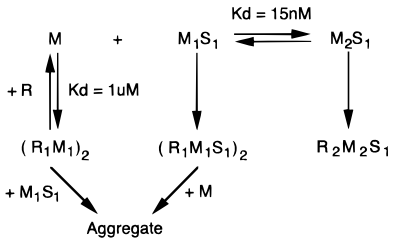

FIGURE 9: Proposed subunit assembly pathway for *Eco*KI nuclease from its constituent subunits with the K_d values for steps which appear reversible in the HPLC gel filtration experiments.

was linearized with both *Eco*RI and *Pvu*II to produce two linear fragments each containing one *Eco*KI site, then very little cleavage by *Eco*KI was observed (Figure 8c), unless a considerable excess of enzyme was present (data not shown). The nuclease produced by *in vitro* assembly of R with mtase had almost identical nuclease activity to the *in vivo* nuclease (data not shown). The cleavage patterns observed were identical, the only difference being a slightly lower rate of cleavage probably due to an overestimate of the active protein concentration. The HPLC elution profiles of the reassembled nuclease show that it is not 100% homogeneous (Figure 7a). In addition, the reassembled nuclease was unable to cleave unmodified plasmid in either circular or linear forms in the absence of SAM, ATP, or Mg^{2+}, cleave *Eco*KI-modified plasmid even in the presence of all cofactors, or cleave DNA containing a single site unless present in excess. Thus, the reassembled nuclease displayed the same cofactor and substrate requirements as the naturally isolated enzyme.

The complex formed between R and M_1S_1 showed very limited cleavage of the DNA substrates (Figure 8d). The rate of cleavage was such that most substrate remains uncleaved after 1 h. We attribute this poor cleavage to the presence of a small amount of M_2S_1 mtase in the M_1S_1 preparation rather than an intrinsic activity of the complex of R and M_1S_1.

The cleavage model described for *Eco*KI (Studier & Bandyopadhay, 1988) was elucidated by analysis of the cleavage of long pieces of phage DNA; however, there have been some reports that circular plasmids only become linearized due to a single cleavage (Weiserova *et al*., 1993). We think this may be due to the use of too low a concentration of active nuclease as we found that circular plasmids were rapidly linearized and then further degraded.

DISCUSSION

We have purified to at least 95% purity all of the subunit components and assembly intermediates, except the insoluble S subunit, of the *Eco*KI type I restriction/modification protein (Figure 1). On mixing the components, we have demonstrated that they can reassociate with each other in several different combinations. Only the M_2S_1 mtase and the $R_2M_2S_1$ bifunctional nuclease/mtase form active enzymes (Figures 2−7). The results of nuclease assays on plasmid substrates agree with earlier studies on the cleavage mechanism of *Eco*KI (Studier & Bandyopadhay, 1988) and show that the *in vitro* assembled nuclease has the same activity as that assembled *in vivo* (Figure 8).

Our results allow us to draw an *in vitro* assembly pathway showing the apparently irreversible assembly of the R subunit with M_1S_1 and M_2S_1 and reversible association of M with M_1S_1 and the R subunit (Figure 9). The K_d for the assembly of the mtase is approximately 15 nM, but the K_d for association of R and M is much weaker, being approximately 1 μM. It is apparent that formation of nuclease is most successful given a preformed mtase or the correct mixture of M and M_1S_1 before adding the R subunit. Mixing of the components in other orders or in different ratios leads to the formation of a large aggregate (Figure 7). The different complexes observed in our assembly experiments appear to be the major species formed, which implies that Figure 9 shows the most important assembly pathways. Our detection methods cannot rule out the existence of other minor pathways.

It is well-known that the estimation of K_d by the application of small sample volumes of different concentrations to gel filtration columns can significantly underestimate the tightness of binding of two components (Ackers, 1975; Gegner & Dahlquist, 1991; Phizicky & Fields, 1995). This is due to dilution and diffusion as the sample moves through the gel. The use of large zone chromatography (Ackers, 1975) or equilibrium methods (Hummel & Dreyer, 1962) can circumvent this problem but requires more sample than is usually available. The small sample volume method which we have used is suitable for complexes which do not dissociate in the concentration range studied and clearly differentiates stable complexes of R with M_1S_1 and M_2S_1 from the unstable *Eco*KI mtase. To limit the effects of zone spreading in our analysis of the dissociation of the mtase, we used the continuous variation titration method in which the protein concentrations are maintained well above the K_d (Job, 1928; Agmus, 1961). The high protein concentration coupled with the very fast elution times obtainable with the HPLC column reduced the effects of zone spreading so that we were able to obtain 75−80% of the theoretical maximum amount of mtase formation. We stress that the K_d for the association of the M subunit with M_1S_1 or the R subunit may still be an underestimate of the tightness of binding.

We believe our results are a reasonably accurate description of the relative strengths of subunit interactions in *Eco*KI, even with the caveats above, and propose that the *in vitro* assembly pathway is relevant to the *in vivo* assembly of a type I R/M enzyme. This proposal allows one to explain, at least partially, the regulation of restriction and modification phenotypes for type I systems and many other experimental observations.

It has long been known that the genes for type I systems and the closely related type III systems can be transferred to a host cell lacking the type I system without the cleavage of the unmodified chromosome that one might expect (Arber & Dussioux, 1962; Boyer, 1964; Colson *et al*., 1965; Fuller-Pace *et al*., 1985; Suri & Bickle, 1985; Loenen *et al*., 1987; Prakash *et al*., 1991; Prakash-Cheng & Ryu, 1993; Prakash-Cheng *et al*., 1993; Redaschi & Bickle, 1996). It has been shown for *Eco*KI that the modification phenotype appears early upon establishment of the system in an unmodified host but that the appearance of the restriction phenotype is long delayed (Prakash-Cheng & Ryu, 1993). The delay in the appearance of the restriction phenotype can be abolished, with lethal effects on the host cell, by a mutation in a gene distant from the chromosomal location of *Eco*KI (Prakash-Cheng *et al*., 1993). They suggested that the primary function of this *hsdC* gene could be different from regulation of R/M phenotype. The two promoters for the chromosomally located *Eco*KI system, one for the *hsdR* gene and

one for both *hsdM* and *S* genes, are known to be very weak and of approximately equal strength, and there is no evidence for regulation of the phenotypes at a transcriptional level (Loenen *et al*., 1987; Prakash *et al*., 1991; Prakash-Cheng *et al*., 1993). The relative intracellular concentrations of the subunits when the *Eco*KI genes are chromosomally encoded are [R] < [S] = 0.5[M] (Weiserova *et al*., 1993). If this balance of subunit concentration is upset by, for instance, raising the gene copy number with a plasmid encoding the *hsdR* gene, then the observed phenotypes can be drastically perturbed (Webb *et al*., 1996). Furthermore, it has been estimated that the *Eco*KI system can be overwhelmed by multiple infection with approximately 60 phage λ (Kelleher & Raleigh, 1994). If one bears in mind that type I enzymes do not turn over in the DNA cleavage reaction and that two target sites are required for efficient cleavage, then these results imply that there are at most 100−200 molecules of *Eco*KI nuclease per cell.

The *in vitro* assembly pathway which we have derived for *Eco*KI (Figure 9) can be used as a mechanism for the posttranslational control of R/M phenotype. It is estimated very roughly that one molecule of a compound in a typical *E. coli* cell is at an effective concentration of about 1−3 nM (Goodsell, 1991). This low concentration is of the same order of magnitude as the K_d for the assembly of the mtase, the essential complex on the pathway to the complete protein. Upon transfer of the *hsd R*, *M*, and *S* genes to an unmodified host, synthesis of all three subunits will commence. Until the concentration of M and S subunits rises substantially above 10 molecules per cell, then they will exist mainly as free M and M_1S_1. The binding of M to R is probably not physiologically relevant due to the poor dissociation constant; however, M_1S_1 will outnumber the molecules of M_2S_1 and be able to bind tightly to most of the available R subunits. These inactive complexes of R and M_1S_1 are a dead end in the pathway and could be removed by the cellular protein degradation mechanisms (Gottesman & Maurizi, 1992). The first few molecules of M_2S_1 that form are statistically unlikely to find any R subunits with which to associate. They can, however, bind tightly to unmodified target sequences on the chromosome and proceed to modify and protect them from the action of the nuclease. Even though this methylation of unmodified DNA is rather inefficient, it should be able to compete with the potential nuclease reaction. It has been found both *in vivo* and *in vitro* that two target sites on a piece of DNA give much more efficient cleavage than DNA with one site (Murray *et al*., 1973; Webb *et al*., 1996). The mtase reaction should be able to compete with the cleavage reaction on the unmodified chromosome since methylation occurs directly at the target sequence while cleavage occurs at a great distance apparently on collision of two translocating type I nucleases (Studier & Bandyopadhay, 1988). Translocation on the chromosome is likely to be very inefficient because of the presence of numerous other proteins bound to the chromosome, but on viral DNA entering the bacterium which will be relatively clear of bound proteins, translocation should be much easier. Supporting this, it has recently been shown *in vitro* that binding of lac repressor protein between two target sites for *Eco*P15I type III restriction endonuclease, which shares some similarity with type I nucleases due to a requirement for ATP-dependent DNA translocation prior to cleavage, can block cleavage by *Eco*P15I (Meisel *et al*., 1995). Therefore, our assembly pathway would allow modification of the chromosome and reduce the potentially deleterious effects of the nuclease. The appearance of the restriction phenotype is delayed until the concentration of M_2S_1 reaches a significant proportion of the concentration of free M and M_1S_1 and is of the order of 10 nM or greater. At this concentration, the assembly of the complete nuclease becomes more probable. The mutation in the *hsdC* gene (Prakash-Cheng *et al*., 1993) which allows early expression of the restriction phenotype with disastrous consequences for the host has not been identified, but on the basis of our assembly pathway, it would appear reasonable to suggest that its absence perturbs the balance of the assembly pathway by allowing a too rapid accumulation of R, M, and S subunits to a level at which M_2S_1 and hence nuclease are more likely to form. HsdC may have a role in degrading the inactive complexes M_1S_1 and $R_1M_1S_1$ or even free R and M subunits. Alternatively, it could bind to unassembled components of *Eco*KI such as the R subunit to prevent their assembly with the rest of the type I system. Such a mechanism has been invoked for the *Pvu*II type II R/M system (Adams & Blumenthal, 1995). We note that although we have written the assembly of R with M_1S_1 and M_2S_1 as irreversible, our measurements are not sensitive enough to detect dissociation at a subnanomolar concentration which could allow the inactive $R_1M_1S_1$ form to be rescued.

The use of a complex subunit assembly pathway to control the R/M phenotype suggests that it should be possible to isolate systems with a mutant phenotype due to defective assembly. Numerous *hsd* mutants of *Eco*KI have been isolated, most of which can be attributed to a loss of DNA binding by S (res^-mod^-), defective restriction by a change in R (res^-mod^+), or altered but still effective modification caused by changes in M ($res^+mod^\pm$). However, there are mutations in the *hsd* genes which have unexpected phenotypic effects and which could be assembly mutants. Indeed, complementation between type I systems which contain different mutations with different phenotypes has been used to define which subunits contain mutations and to which family a system belongs (Modrich, 1979; Yuan, 1981; Barcus & Murray, 1995), indicating that subunit exchange can occur *in vivo* during the assembly of the nuclease. Several temperature-sensitive mutations in the *hsdS* gene have been found which cause a temperature-dependent loss of mtase activity. It was proposed that this was due to a reduction in the stability of the M_2S_1 mtase rather than a direct effect on catalysis (Zinkevich *et al*., 1990, 1992). This effect has recently been confirmed (unpublished results, D. Dryden, K. Firman, V. Zinkevich, and B. Hussey). We have found that the equilibrium between the mtase and the M and M_1S_1 forms is shifted in favor of the partially assembled forms in a temperature-dependent manner.

The proposed assembly pathway for *Eco*KI is very likely applicable to all other type I R/M enzymes. *Eco*BI, the only other type IA enzyme to have been studied in detail, was purified as a mixture of protein species with different proportions of R and M subunits and one S subunit (Eskin & Linn, 1972; Lautenberger & Linn, 1972). During storage of the protein, the stoichiometry of the protein continued to change. *Eco*AI, the only purified member of the type IB family (Suri *et al*., 1984), was found to dissociate during chromatography into a fraction containing M and S subunits which was active as an mtase and a fraction containing the R subunit. Mixing these two fractions restored nuclease

activity. This can be accommodated in our assembly scheme by making the association of two R subunits with the mtase a reversible equilibrium reaction. It was not noted if the *Eco*AI mtase would dissociate. Our preliminary information about the stability of the mtase part of the most well-characterized type IC enzyme, *Eco*R124I, indicates that it does not dissociate at nanomolar concentrations (unpublished results, D. Dryden, K. Firman, and P. Janscak). Therefore, *Eco*R124I appears to have an irreversible association of M and M_1S_1 to form the mtase rather than an equilibrium situation as found for *Eco*KI. It appears that although the assembly of the mtase is irreversible, the subsequent assembly with the R subunit is reversible as the nuclease can dissociate during ion exchange chromatography (Janscak *et al*., 1996). It has been found for both *Eco*AI and *Eco*R124I that transfer of the genes to a new nonrestricting host is difficult if the host already contains the R subunit or can express more of the R subunit than of the M and S subunits (Suri & Bickle, 1985; Fuller-Pace *et al*., 1985; Janscak *et al*., 1996), suggesting that the cellular concentration of the R, M, and S subunits for these sytems is as tightly controlled as for the *Eco*KI system. Therefore, we suggest that by changing the balance of the irreversible subunit associations and the reversible subunit associations, it is possible to account for many of the phenotypic and biochemical observations relating to the competition between restriction and modification by these systems.

The complexity of type I systems is such that it is difficult to understand why they have been maintained by bacteria when the much simpler type II R/M systems can provide adequate protection against foreign DNA. The ease with which DNA sequence specificity can be changed by altering only the *hsdS* gene which changes the target specificity for both restriction and modification simultaneously is one potential advantage over type II systems (King & Murray, 1994). A second advantage may be the possibilities for subunit complementation and the establishment of new type I systems in unmodified hosts that is offered by the assembly pathway presented above. This may be particularly relevant for the maintenance of a diverse population of these systems and to the transfer of type I systems between species (Sharp *et al*., 1992; Murray *et al*., 1993).

ACKNOWLEDGMENT

It is a pleasure to thank Professor N. E. Murray and our colleagues in her laboratory for their support and advice. We also thank M. Barrett and F. Paulin for their assistance in early experiments, N. Mistry for technical assistance with the ultracentrifuge experiments, and the WELMET amino acid sequencing service at Edinburgh University.

REFERENCES

Ackers, G. K. (1975) in *The Proteins* (Neurath, H., Hill, R L., & Boeder, C.-L., Eds.) 3rd ed., pp 1−94, Academic Press, New York.

Adams, G. M., & Blumenthal, R. M. (1995) *Gene 157*, 193−199.

Agmus, E. (1961) *Fresenius' Z. Anal. Chem. 183*, 321−333.

Barcus, V. A., & Murray, N. E. (1995) in *Population Genetics of Bacteria* (Baumberg, S., Young, J. P. W., Saunders, S. R., & Saunders, E. M. H., Eds.) Society for General Microbiology Symposium 52, pp 31−58, Cambridge University Press, New York.

Barcus, V. A., Titheradge, A. J. B., & Murray, N. E. (1995) *Genetics 140*, 1187−1197.

Bickle, T. A. (1993) in *Nucleases* (Linn, S. M., Lloyd, R. S., & Roberts, R. J., Eds.) 2nd ed., pp 80−109, Cold Spring Harbor Press, Cold Spring Harbor, NY.

Bickle, T. A., & Kruger, D. H. (1993) *Microbiol. Rev. 57*, 434−450.

Buhler, R., & Yuan, R. (1978) *J. Biol. Chem. 253*, 6756−6760.

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J.-F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M., Weidman, J. F., Fuhrmann, J. L., Nguyen, D., Utterback, T. R., Kelley, J. M., Peterson, J. D., Sadow, P. W., Hanna, M. C., Cotton, M. D., Roberts, K. M., Hurst, M. A., Kaine, B. P., Borodovsky, M., Klenk, H.-P., Fraser, C. M., Smith, H. O., Woese, C. R., & Venter, J. C. (1996) *Science 273*, 1058−1073.

Burckhardt, J., Weisemann, J., Hamilton, D. L., & Yuan, R. (1981) *J. Mol. Biol. 153*, 425−440.

Clewell, D., & Helinski, D. R. (1969) *Proc. Natl. Acad. Sci. U.S.A. 62*, 1159−1166.

Cooper, L. P., & Dryden, D. T. F. (1994) *J. Mol. Biol. 236*, 1011−1021.

Cowan, G. M., Gann, A. A. F., & Murray, N. E. (1989) *Cell 56*, 103−109.

Creeth, J. M., & Harding, S. E. (1982) *J. Biochem. Biophys. Methods 7*, 25−34.

Daniel, A. S., Fuller-Pace, F. V., Legge, D. M., & Murray, N. E. (1988) *J. Bacteriol. 170*, 1775−1782.

Dryden, D. T. F., Cooper, L. P., & Murray, N. E. (1993) *J. Biol. Chem. 268*, 13228−13236.

Dryden, D. T. F., Sturrock, S. S., & Winter, M. (1995) *Nat. Struct. Biol. 2*, 632−635.

Dybvig, K., & Yu, H. (1994) *Mol. Microbiol. 12*, 547−560.

Eskin, B., & Linn, S. (1972) *J. Biol. Chem. 247*, 6183−6191.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J. D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. F., Smith, H. O., & Venter, J. C. (1995) *Science 269*, 496−512.

Fuller-Pace, F. V., Cowan, G. M., & Murray, N. E. (1985) *J. Mol. Biol. 186*, 65−75.

Gann, A. A. F., Campbell, A. J. B., Collins, J. F., Coulson, A. F. W., & Murray, N. E. (1987) *Mol. Microbiol. 1*, 13−22.

Gegner, J. A., & Dahlquist, F. W. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 750−754.

Goodsell, D. S. (1991) *Trends Biochem. Sci. 16*, 203−206.

Gorbalenya, A. E., & Koonin, E. V. (1991) *FEBS Lett. 291*, 277−281.

Gottesman, S., & Maurizi, M. R. (1992) *Microbiol. Rev. 56*, 592−621.

Hadi, S. M., & Yuan, R. (1974) *J. Biol. Chem. 249*, 4580−4586.

Harding, S. E., Horton, J. C., & Morgan, P. J. (1992) in *Analytical ultracentrifugation in biochemistry and polymer science* (Harding, S. E., Rowe, A. J., & Horton, J. C., Eds.) pp 275−294, Royal Society of Chemistry, Herts, U.K.

Hummel, J. P., & Dreyer, W. J. (1962) *Biochim. Biophys. Acta 63*, 530−532.

Janscak, P., Abadjieva, A., & Firman, K. (1996) *J. Mol. Biol. 257*, 977−991.

Job, P. (1928) *Ann. Chim. (Ser. 10) 9*, 113−134.

Kelleher, J. E., & Raleigh, E. A. (1994) *J. Bacteriol. 176*, 5888−5896.

Kelleher, J. E., Daniel, A. S., & Murray, N. E. (1991) *J. Mol. Biol. 221*, 431−440.

King, G., & Murray, N. E. (1994) *Trends Microbiol. 2*, 465−469.

Klotz, I. M., Darnall, D. W., & Langerman, N. R. (1975) in *The Proteins* (Neurath, H., Hill, R. L., & Boeder, C.-L. Eds) 3rd ed., pp 293−411, Academic Press, New York.

Kneale, G. G. (1994) *J. Mol. Biol. 243*, 1−5.

Lautenberger, J. A., & Linn, S. (1972) *J. Biol. Chem. 247*, 6176−6182.
Loenen, W. A. M., Daniel, A. S., Braymer, H. D., & Murray, N. E. (1987) *J. Mol. Biol. 198*, 159−170.
Meisel, A., Mackeldanz, P., Bickle, T. A., Kruger, D. H., & Schroeder, C. (1995) *EMBO J. 14*, 2958−2966.
Meselson, M., Yuan, R., & Heywood, J. (1972) *Annu. Rev. Biochem. 41*, 447−466.
Modrich, P. (1979) *Q. Rev. Biophys. 12*, 316−369.
Murray, N. E., Batten, P. L., & Murray, K. (1973) *J. Mol. Biol. 81*, 395−407.
Murray, N. E., Gough, J. A., Suri, B., & Bickle, T. A. (1982) *EMBO J. 1*, 535−539.
Murray, N. E., Daniel, A. S., Cowan, G. M., & Sharp, P. M. (1993) *Mol. Microbiol. 9*, 133−143.
Nagaraja, V., Shepherd, J. C. W., Pripfl, T., & Bickle, T. A. (1985) *J. Mol. Biol. 182*, 579−587.
Phizicky, E. M., & Fields, S. (1995) *Microbiol. Rev. 59*, 94−123.
Powell, L. M., Dryden, D. T. F., Willcock, D. F., Pain, R. H., & Murray, N. E. (1993) *J. Mol. Biol. 234*, 60−71.
Prakash, A., Valinluck, B., & Ryu, J. (1991) *Gene 99*, 9−14.
Prakash-Cheng, A., & Ryu, J. (1993) *J. Bacteriol. 175*, 4905−4906.
Prakash-Cheng, A., Chung, S. S., & Ryu, J. (1993) *Mol. Gen. Genet. 241*, 491−496.
Redaschi, N., & Bickle, T. A. (1996) *J. Mol. Biol. 257*, 790−803.
Sain, B., & Murray, N. E. (1980) *Mol. Gen. Genet. 180*, 35−46.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A laboratory manual*, Cold Spring Harbor Press, Cold Spring Harbor, NY.
Sharp, P. M., Kelleher, J. E., Daniel, A. S., Cowan, G. M., & Murray, N. E. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 9836−9840.
Sober, H. A. (1970) *Handbook of Biochemistry*, 2nd ed., pp B75−B76, CRC Press, Boca Raton, FL.
Srivastava, D. K., & Bernhard, S. A. (1987) *Annu. Rev. Biophys. Biophys. Chem. 16*, 175−204.
Stein, D. C., Gunn, J. S., Radlinska, M., & Piekarowicz, A. (1995) *Gene 157*, 19−22.
Studier, F. W., & Bandyopadhay, P. K. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4677−4681.
Suri, B., & Bickle, T. A. (1985) *J. Mol. Biol. 186*, 77−85.
Suri, B., Shepherd, J. C. W., & Bickle, T. A. (1984) *EMBO J. 3*, 575−579.
Taylor, I., Watts, D., & Kneale, G. (1993) *Nucleic Acids Res. 21*, 4929−4935.
Valinluck, B., Lee, N. S., & Ryu, J. (1995) *Gene 167*, 59−62.
Webb, J. L., King, G., Ternent, D., Titheradge, A. J. B., & Murray, N. E. (1996) *EMBO J. 15*, 2003−2009.
Weiserova, M., Janscak, P., Benada, O., Hubacek, J., Zinkevich, V. E., Glover, S. W., & Firman, K. (1993) *Nucleic Acids Res. 21*, 373−379.
Xu, G., Willert, J., Kapfer, W., & Trautner, T. A. (1995) *Gene 157*, 59.
Yuan, R. (1981) *Annu. Rev. Biochem. 50*, 285−315.
Zinkevich, V. E., Weiserova, M., Kryukov, V. M., & Hubacek, J. (1990) *Gene 90*, 125−128.
Zinkevich, V. E., Heslop, P., Glover, S. W., Weiserova, M., Hubacek, J., & Firman, K. (1992) *J. Mol. Biol. 227*, 597−601.

BI9619435