Published in final edited form as:
J Mol Biol. 2007 February 23; 366(3): 768–778.

A Type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs.

Catherine L Bair[#] and **Lindsay W Black**[%]
Department of Biochemistry and Molecular Biology, University of Maryland School of Medicine, Baltimore, MD 21201-1503, USA

Abstract

The *E. coli* CT596 prophage exclusion genes *gmrS* and *gmrD* were found to encode a novel Type IV modification dependent restriction nuclease that targets and digests glucosylated (glc)-hydroxymethylcytosine (HMC) DNAs. The protein products GmrS (36 kDa) and GmrD (27 KDa) were purified and found to be inactive separately, but together degraded several different glc-HMC modified DNAs (T4, T2, T6). The GMR enzyme is able to degrade both α–glucosy-HMC T4 DNA and β–glucosyl-HMC T4 DNA, whereas no activity was observed against non-modified DNAs including unmodified T4 cytosine (C) DNA or non-glucosylated T4 HMC DNA. Enzyme activity requires NTP, favors UTP, is stimulated by calcium, and initially produces 4 kb DNA fragments that are further degraded to low molecular weight products. The enzyme is inhibited by the T4 phage Internal Protein I* (IPI*) to which it was found to bind. Overall activities of the purified GmrSD enzyme are in good agreement with the properties of the cloned *gmr* genes *in vivo* and suggest a restriction enzyme specific for sugar modified HMC DNAs. IPI* thus represents a third generation bacteriophage defense against restriction nucleases of the Gmr type.

Keywords

restriction; DNA modifications; DNA injection; virus evolution; *Myoviridae*

Introduction

The terms exclusion and restriction both describe the abortive infection of a host by a bacteriophage. Exclusion, as it is used in this work, is a more general term for the inability of a phage to successfully infect a bacteria, whereas restriction is used to describe specific (site- and base-specific) endonuclease digestion of an invading phage genome (reviewed in reference1). Restricted phage DNA is targeted either by its lack of a host-like methylation pattern or by the presence of unusual base modifications relative to the host DNA (e.g. hydroxymethylcytosine (HMC) residues). Restriction modification (R-M) system and modification-dependent systems (MDS) are the terms used to describe two categories of nuclease-based host defense systems that target foreign DNAs. Sequence specific R-M systems appear to be more prevalent, whereas MDS's are a more sparsely populated subset whose activities are generally directed more towards a modified base rather than just a defined

[%] Corresponding author.
[#]Present address: NIH/NCI Laboratory of Molecular Biology Develeopmental Genetics Section, 37 Convent Dr. Bethesda, MD 20892.

sequence. For MDS's, the initiation of digestion occurs only after the enzyme encounters a specific modification present on only the invading DNA2.

These nuclease-based systems are loosely grouped into four main Types (I–IV) based upon a number of criteria3, however the classification may change as an enzyme is better characterized. R-M systems of Type I, Type II and Type III are typically paired methyltransferase and endonuclease activities. These two activities act in concert to protect the host genome through methylation and, inversely, digest foreign DNA that lacks this methylation pattern 2,4,5. The majority of *E. coli* strains possess Type I R-M systems, whereas the Type III class of R-M systems is the least populated 6,7. Type IV, although it contains some methylation-based systems, chiefly contains MDS's that target foreign DNA based upon its unusual base modifications8,9. In addition to a Type I enzyme, *E. coli* K12 contains three MDS systems (*Mcr*A, *Mrr*, and *Mcr*BC10,11) that specifically recognize modified DNA. Among these, McrBC, which targets Teven HMC DNAs, is the best characterized (see *Discussion*).

An accompanying manuscript1 describes the exclusion properties of the *E. coli* CT596 genes *gmrS* and *gmrD* that specifically block infection of Teven phages lacking a functional *ip1* gene. Presented in this work are the purification and biochemical characterization of the GmrS and GmrD proteins that are shown together to form a modification dependent restriction enzyme, GmrSD, that can digest Teven glc-HMC DNAs in the presence of NTP hydrolysis. Until now, no restriction enzyme had been described that specifically targets the sugar-modified HMC residues of Teven phage family members to inhibit successful host infection. The GmrSD enzyme is able to degrade a number of different glucosylated HMC DNAs, including α–glucosyl-, β–glucosyl, and gentiobiosyl-HMC DNAs, but displays no detectable activity against non-modified DNAs or against HMC T4, the preferred McrBC substrate. Consistent with the role of the Teven *ip1* locus genes in infection of *E. coli* CT596 and its *gmrS/gmrD* clones, the T4 IPI* protein is observed *in vitro* to block GmrSD nuclease activity and interact with the GmrS and GmrD proteins.

Results

Purification of GmrS and GmrD

A number of protein tag-based expression and purification vectors were used (His-tag, HSV tag, ion exchange and GST tag) without success to purify the GmrS and GmrD proteins individually or together due to apparent insolubility, failure to remove tags by proteolysis and apparent proteolytic degradation of GmrS in the presence of GmrD. The best results came from using the self-cleaving chitin binding domain (CBD)/intein tag as a purification vehicle (IMPACT-CN, New England Biolabs), as well as processing the two proteins separately. A typical purification of the GMRs and GMRd proteins is shown in Figure 1. As seen in Figure 1, both GmrS (~36 KDa) and GmrD (~27 KDa) co-purified with *Gro*EL, the major eluate from the chitin column, which may exhibit affinity to chitin. Western Blot analyses were performed on duplicate gels to verify the identity of the chaperonin *Gro*EL and residual chitin binding domain proteins that are identified on Figure 1. Attempts to remove *Gro*EL from the chitin resin were not successful in the presence of GmrS or GmrD yet it was nearly completely removed from a Blank Vector preparation, suggesting that most *Gro*EL binding is due to affinity for the unfolded CBD-GmrX fusion proteins.

The identities of the GmrS and GmrD proteins purified by this chitin column procedure were confirmed by mass spectrometry of the proteins eluted from gel slices following SDS-PAGE as shown in Figure 1. Tryptic digests of the two proteins were computer matched with high probability (scores 527 GmrS, 282 GmrD) to their DNA sequences, GenBank accession nos.

213227769 and **21327771**, confirming that these two major chitin column eluates are indeed the products of the cloned and expressed *gmrS* and *gmrD* genes.

Attempts to purify further active GmrS and GmrD eluted from chitin columns were unsuccessful. Both GmrS and GmrD displayed high affinity for the *Gro*EL protein, and co-eluted with *Gro*EL upon FPLC Superose 6 size exclusion column chromatography (just ahead of the thyroglobulin 670 kDa size marker), even in the presence of high salt and ATP (1 M NaCl, 10 mM $MgCl_2$, 5 mM ATP) (data not shown). It was presumably because of this high affinity that other purification steps following chitin affinity purification were also unsuccessful. Likely, *Gro*EL aids in the solubility and proper folding of GmrS and GmrD and may slowly release the two proteins to interact and form active enzyme upon mixing. Since it was not possible to remove the *Gro*EL protein, purification of *Gro*EL protein through its adsorption to chitin was performed in parallel and used as a control (the Blank Control Vector) in enzymatic characterization experiments, together with the individual GmrS and GmrD proteins as controls.

Nuclease activity of GmrS and GmrD

Initial nuclease assays were performed on DNA substrates from burst frozen/thawed phage after five cycles of incubation in dry ice-ethanol and water at 37 °C. This procedure was used in order to release encapsidated DNA still bound to the internal phage proteins, including IPI*. Approximately 300 ng of each of the purified GmrS and GmrD proteins were used in addition to 0.5 mM UTP, and as shown later, both proteins were required for activity. Figure 2 shows the digestion of the DNAs released from glc-HMC phages T4, T2, T6, and T4 *ip1*$^-$ mutants eG506 and HA35 by the GmrSD enzyme, where only wtT4 DNA (lane 1) was protected due to the presence of its IPI* protein. The agarose gel electrophoresis showed initial digestion of the T2, T6 and T4 *ip1*$^-$ mutants' DNAs that are not protected by the IPI* protein into a discrete band at ~4 kb and a smear of other larger and smaller fragments (lanes 2–5). The small amount of digestion of wtT4 DNA may be due to imperfect IPI* protection. Overall, the assay showed a strong *in vitro* activity directed towards Teven phage glc-HMC DNAs not protected by the T4 IPI* protein.

Further experimentation was performed on protease treated, phenol extracted, pure phage DNAs in order to determine the protein and ionic requirements for optimal GmrSD activity. The results (data not shown) revealed that GmrS and GmrD appeared to act preferentially at a molar ratio of 3–4 GmrD:1 GmrD, and activity was optimal in 5 mM $CaCl_2$, 2 mM KOAc, 3 mM $MgCl_2$ and 1 mM UTP on glc-HMC phage DNA substrates. The degradation of phage DNAs assessed on agarose gels was not apparent until 15 min of digestion, with a visually assessed virtual 'limit' digestion generally seen as the disappearance of the 170 kb genome band after 35min at 37 °C. Under low activity conditions a ~4 kb intermediate comparable to that seen in Figure 2 was also observed in the phenol extracted DNA, showing that this intermediate is not a consequence of the packaged DNA structure (data not shown).

Specificity of the GmrS and GmrD enzyme

Control assays were performed to verify that the digestion of Teven DNA was not the result of background nonspecific nuclease activity. Non-modified plasmid DNA (covalently closed circular pBR322) was neither nicked nor cleaved in the presence of GmrS or GmrD individually or together, or in the presence of the *Gro*EL control preparation (data not shown). The GmrSD enzyme did not hydrolyze *E. coli* CT596 DNA (data not shown). Moreover, the GmrSD enzyme did not hydrolyze T4 HMC DNA or T4 C DNA, nor did it degrade T7 DNA (Figure 3A (lanes 1–7)). Taken together, these assays therefore demonstrated a nuclease sufficiently pure for further enzymatic characterization, as well as one that mimicked *in vitro* the *in vivo* restriction properties of the *gmrS*/*gmrD* genes.

NTP requirement for the GmrSD enzyme

Nuclease assays performed under optimal conditions established that addition of NTP is absolutely required for GmrSD enzyme activity on glc-HMC DNAs. Although there is a preference for UTP, good activity is also observed with GTP or CTP, thus the enzyme is likely an NTPase that prefers UTP but has the ability to utilize GTP for activity (Figure 3B (lanes 1–6)). Evidence that nuclease activity requires hydrolysis of NTP is provided by high activity in the presence of GTP but not of GMP or GDP, and by inhibition of activity by GTP-γ-S (lanes 7–9). Of these, only GTP-γ-S inhibited activity upon addition to UTP, beginning from a ratio of 0.25 GTP-γ-S:1UTP.

Further evidence for GmrSD UTPase activity was provided by assays of UTP conversion to UDP (UTP α–P33) measured by PEI cellulose chromatography on concentrated Superose 6 FPLC chromatographed GmrS and GmrD proteins. GmrS showed UTPase activity (higher on a molar basis than control *Rec*A protein assayed with DNA under GmrSD nuclease conditions12), whereas GmrD did not show activity (data not shown). These data support the identification of IBEGs as a UTPase. However, conclusive data await further enzymatic and mutational analysis, since GmrD and DNA were not stimulatory individually or together under our assay conditions.

Nuclease substrate specificity

Nuclease assays were performed on wtT4 DNA, T4 DNA containing only α-glcHMC, T4 DNA containing only β-glcHMC, and other Teven phage DNAs. As shown in Figure 4, (lanes 4, 6, 8, 10, 12 and 13) the GmrSD enzyme digests all glc-HMC substrates; and this activity requires UTP. In agreement with the *in vivo* studies of transformants expressing only one of the *gmr* genes, designated SI and ID1 that lacked restriction of Teven phages, neither GmrS nor GmrD was active on any of these phage DNA substrates in isolation (lanes 7, 9, 11, 13, 14 and 15). Digestion occurred regardless of the type of bond between the glucose moiety and the HMC residue since wtT4 (70%α-glc, 30% β-glcHMC) DNA, T2 [70% α-glc, 5% gentiobiosyl-HMC) DNA,and T6 (3% α-glc, 70% gentiobiosyl-HMC) DNAs all yielded positive “limit” DNA digests following nuclease digestion. Thus from the T6 digestion result the diglucose (gentiobiose) HMC modification which predominates in this DNA is apparently susceptible to digestion13,14. However, wtT4 DNA appeared to be preferentially digested as compared to T2 and T6 DNAs (lane 2, T2 and lane 10, T6) in comparison to the complete loss of T4 DNA in lanes 4 and 12). Moreover, the nuclease also appeared to favor digestion of α-glc HMC T4 DNA to β-glc T4 DNA (lanes 6 and 8, respectively). In the absence of T4 β-glucosyltransferase, α-glucosyltransferase is able to modify only 75% of the DNA, whereas β-glucosyltransferase functions to modify all HMC sites when α-glucosyltransferase is absent. Thus, the β-glcHMC substrate should have more target residues but appears less well digested15. However, a more precise conclusion about relative glc-HMC susceptibilities to the GmrSD enzyme will require kinetic studies of simpler, model sugar HMC oligonucleotide substrates.

Mutant enzyme studies

Although all phages lacking IPI* were restricted by *gmrS/gmrD* clones, phages containing the IPI* and IP5* proteins displayed different specificities *in vivo* toward a clone expressing both native GmrS and a GmrD protein that is fused at its C-terminal end to a 6XHis tag (clone L7H). The presence of this tag on GmrD did not affect the ability of the purified GmrSD enzyme to digest genomic DNAs *in vitro* since known substrates show the same digestion when the chitin column purified GmrD-6XHis protein (see L7H purification in **Methods**) is combined with the GmrS protein (data not shown). The observed greater sensitivity of this clone to certain phages could be explained in several ways. One possibility is that the GmrSD-6XHis enzyme has reduced activity, allowing for certain phages to repair their genomes faster than the enzyme

can inactivate them. Alternatively, the C-terminal run of histidine residues may result in increased sensitivity of GmrSD-6XHis to the IP5* protein to which it is otherwise relatively immune.

TaqI DNA fragment binding and digestion studies

DNA binding of the GmrS and GmrD proteins was studied by DNA PAGE of a DIG (Roche)-labeled specific ~1 kb *Taq*I restriction endonuclease fragment of wt T4 DNA (*Taq*I is unaffected by T4 DNA modifications). Electrophoresis of the protein bound DNA fragment suggested preferential binding to HMC modified DNA, since control cDNA did not inhibit binding whereas T4 DNA did. Moreover, addition of IPI* protein prevented DNA binding (data not shown). To investigate the effect of DNA sequence and length on GmrSD endonuclease activity, other specific *Taq*I fragments of T4 DNA less than 10 kb in length were gel purified and subjected to digestion. The GmrSD enzyme was able to digest most gel-purified unique wtT4 *Taq*I fragments tested, with the notable exception of two: a 4.5 kb and a ~2.0/2.5 kb species (Figure 5). Failure of the GmrSD enzyme to digest these fragments suggests a requirement for a combination of length, specific sequence(s) and/or modification(s) to initiate cleavage. These requirements might be connected to the ~4 kb DNA fragments that appear to accumulate as possible early intermediates, here under lower enzyme activity conditions as described in Methods (Figure 2 and data not shown). Since the enzyme eventually digests all of the T4 DNA to fragments less than 0.5 kb and it digests most of the specific *Taq*I DNA fragments to less than 0.5 kb fragments, there must be a relaxation of the observed specificity requirements following initiation of hydrolysis (See discussion).

Protein:Protein Associations

Protein associations between GmrS, GmrD and IPI* were investigated by several techniques (native polyacrylamide gel electrophoresis, size fractionation on a Superose 6 column, immunoprecipitation, and affinity chromatography). A specific GmrS and GmrD complex could not be demonstrated due to the *Gro*EL contamination, however co-immunoprecipitation assays demonstrated association of the GmrS and GmrD proteins with the IPI* phage inhibitor. Wild type T4 phage or T4 *ip1*$^{-}$ phage were used as the positive and negative sources of IPI*, respectively. GmrS and GmrD were combined with burst wt T4 phage or T4 *ip1*$^{-}$ phage, and with gamma-globulin from an antiserum against IPI*. Incubation with protein A Sepharose beads, to which the IgG binds, followed by centrifugation isolates IPI* bound (IB) proteins from the unbound (UB) proteins.

In the presence of burst wt T4, there appeared to be a strong association between IPI* and GmrS and GmrD. As seen in Figure 6 (lanes 4 and 3) and (lanes 6 and 5), GmrS and GmrD are pulled down by IPI* following antibody reaction (IB) and therefore removed from the supernatant (UB). When mixed, both Gmr proteins co-immunoprecipitate with IPI* (lane 8 versus 7). To validate the specificity of the above results, the experiment was repeated using an equal quantity of T4 *ip1*$^{-}$ phage, which releases ~700 molecules of other basic internal proteins per phage particle16, and which shows that there is no nonspecific association with these other internal proteins and the Gmr proteins. Neither the GmrS nor GmrD protein was precipitated from the solution by an association with the IPI* antibody in the absence of IPI*, since both proteins remain in the supernatant (UB) (lane 9 versus 10). The bottom panels show the immunoprecipitated IPI* that runs as a doublet present in the lower left hand (wt T4) panels but not present in the right hand (T4 IPI^{-}) panels. Comparable results were obtained in chromatography studies using purified nickel column-bound His-IPI* protein, although binding of IPI*to GmrD appeared to be weaker than to GmrS (data not shown). Overall, it appears that IPI* inhibition of GmrSD enzyme activity reflects protein-protein interaction with one or both of these proteins. This interaction is not dependent upon GmrS, GmrD nor IPI* association with DNA, since the co-immunoprecipitation experiments gave comparable results

in the presence (burst phage) or absence (purified IPI*) of T4 DNA. These observations are consistent with the *in vivo* observation that the IPI* inhibition of GmrSD restriction can occur *in trans* from expression vectors where IPI* is uncoupled from DNA transfer from the phage particle.

Discussion

In this work it is shown that the GmrSD enzyme is a novel Type IV modification dependent restriction endonuclease. Nuclease activity is dependent upon the presence of both the GmrS and the GmrD proteins. The enzyme is active on α-glc HMC, β–glc HMC, and gentiobiosyl-HMC DNAs and apparently can also attack the non-glucose, possibly mannose modified, HMC DNA of phage RB69 (data not shown) (Karam, JD, personal communication). Initial digestion of this phage DNA by GmrSD produces fragments of approximately 4 kb. The enzyme was inactive on non-glucosylated HMC containing DNA and non-modified cytosine DNAs. Several features of the enzyme are unusual: i) the enzyme requires UTP, GTP or CTP, and is inhibitied by ATP when at a five-fold excess compared to UTP; ii) apparently requires hydrolysis for activity since it is inactive with GMP, GDP and the non-hydrolyzable GTP–γ S analog; and iii) requires or is greatly stimulated by Ca^{+2}, which is rare for a Type IV restriction enzyme. From these properties, it can be speculated that the enzyme is among the multimeric DNA translocating restriction endonucleases17, however mechanistic studies to establish this and some other enzyme properties require a homogeneous (*Gro*EL-free) enzyme preparation. For our studies, the nuclease activity was judged highly purified based upon the absence of nicking and cutting activities on unmodified DNAs by the *Gro*EL contaminated GmrS and GmrD protein preparations.

Only some purified T4 *Taq*I restriction fragments were hydrolyzed by the GmrSD enzyme, and initial digestion of genomic DNA produces a broad spectrum of high molecular weight fragments from T4, T2, T6 and other DNAs, under some conditions accumulating intermediate ~4 kb fragments, which eventually disappear to low molecular weight fragments. This suggests that cleavage is likely initiated on a subset of available glc-HMC residues determined by sequence or glc HMC-glc-HMC spacing in the DNA. Likely mechanisms include: 1) translocation by 4 kb of the enzyme from an initial recognition-binding site to a cleavage site; 2) looping of DNA between two interacting enzymes bound to individual glc-HMC residues followed by cleavage; or 3) a requirement for 4 kb intervals between enzyme binding events. Subsequent to the initial 4 kb cleavage, secondary step-wise digestion of the initial product would follow, possibly by NTP driven translocation coupled to hydrolysis8,19. Comparable mechanisms have been proposed for DNA translocating restriction endonucleases including McrBC17,19,21.

Several of the enzymatic properties of the GmrSD enzyme resemble those of the thoroughly characterized *Mcr*BC enzyme. In fact, high *Gro*EL affinity was also seen in the purification of the *Mcr*BC restriction enzyme, although *Gro*EL could be removed18. The *Mcr*BC enzyme complex utilizes GTP-driven DNA translocation8,19 and a length-determining looping mechanism to create non-specific fragments of DNA of a characteristic length through multiple cuts in a small region around the HMC recognition site in substrate DNA. These larger DNA intermediates are eventually digested to small (<500bp) fragments from HMC phage DNAs8,20,23. Interestingly, inhibition of the *Mcr*BC enzyme also occurred when ATP was at five times the concentration of GTP8.

The GmrSD enzyme also demonstrates several genetic similarities to the *Mcr*A and *Mcr*BC enzymes. Each of these MDRs is apparently encoded by cryptic prophage-like elements *in E. coli* 10,1. Each enzyme is apparently directed against members of the *Myoviridae* family that contain the HMC (McrBC) or glc-HMC (GmrSD) modified DNA. The *Mcr*BC enzyme is

composed of two enzymatic polypeptides and a third regulatory peptide that are all synthesized from two genes10,22,24, whereas enzymatic studies reported here, as well as gene truncation studies1, show that GmrSD has only two subunits essential for activity. Chief among their differences, the GmrSD enzyme specifically targets sugar modified HMC containing DNA while the *Mcr* enzyme is blocked by this further modification of HMC residues. In this respect the GmrSD enzyme displays some similarities to the *Pvu*Rts1I enzyme that arises from the Rts1 kanamycin resistance plasmid and has been shown to preferentially hydrolyze glc-HMC DNAs to HMC DNAs in certain substrates. However the degree of its specificity and other properties appear distinct from the GmrSD enzyme31.

Overall, it appears that the GmrSD enzyme represents an evolutionary step beyond the *Mcr*A and *Mcr*BC enzymes that occurred in response to the additional DNA modification of the *Myoviridae* family to HMC DNA10. Following upon the likely sequential phage DNA modifications mC, HMC, and a diverse set of glc-HMC modifications, some of the latter targeted by the GmrSD enzyme, these phages have responded by encoding a third generation response, injected protein inhibitors such as IPI* that counters the GmrSD but not the *Mcr* enzymes (Figure 7). Clearly IPI* interacts with the Gmr proteins *in vitro* and *in vivo*1, and it is likely that the other *ip1* locus proteins target other, possibly Gmr-related MDR enzymes. Whether *ip1* locus protein specificity of inhibition is determined solely by binding specificity to Gmr family enzymes or whether the polymorphic sugar-HMC DNA targets also determine specificity remains to be determined.

Materials and Methods

Phage

T4 *ip1* mutants, phage T7, and a number of pseudo- and schizo Tevens were described previously1. T4GT7 (*am*E51g56, *am*C87g42, rNB5060, *alc*) allows for the preparation of cytosine T4 DNA25. Glucosyltransferase mutants T4129 (R20αgt, *am*β*gt*40); T4130 (*am*α*gt*16, β*gt*20); T4106 (α*gt*1); and T4147 (α*gt*1, β*gt*7) were kindly provided by Dr. L. Snyder, Michigan State University and allows for the preparation of α-glucosyl- HMC T4 DNA, β-glucosyl-HMC T4, and HMC T4 DNAs26.

E.coli strains

CT596 is a clinical isolate of *E. coli* serotype K121, DH10B1, BL21 (DE3)1, W3110 (lambda$^-$, r^+, rgl^+), W4597 (W3110 UDPG Ppase$^-$ derivative: *gal*U$^-$, r^+ rgl^+)26, K803 (r_k^-, m_k^-, rgl^-, *sup*E44) kindly provided by Mary Berlyn (E. coli Genetic Stock Center of Yale University), CR63r$^-$r$^-$ (r_2^-, $r_{4,6}^-$, *rgl*B$^-$, *sup*D), B40 (r^+, rgl^+, *sup*D$^+$), CR63λ (lambda$^+$, *sup*D), P301 (r^+, *sup*$^-$, *arg*$^-$), B834*gal*U56 (B834 derivative: *gal*U$^-$, r^-, rgl^+)26.

Vectors

The pTYB2 plasmid for chitin binding purification was purchased from NEB. This method of purification was chosen for its mild operating conditions, tolerance of TritonX-100 (0.1% necessary for solubility of both GmrS and GmrD), mild chemical (DTT) removal of the intein/chitin binding domain tag (the intein undergoes self-cleavage from the protein of interest in reducing or low pH environments) and a lack of high background non-specific binding of numerous *E. coli* proteins to the resin.

Subclone constructs

One construct used for mutant enzyme studies contains the *gmrD*-6XHis (*cac*) sequence downstream of a native *gmrS* sequence that was ligated into pBeloBAC11 to yield plasmid pL7H. This construct was used for *in vivo* studies of the restriction profile of GmrS and GmrD

with the altered 6XHis-tag C-terminal end of GmrD. Another construct contains this His-tag GmrD cloned into the pTYB2 vector to yield a protein fused at its C-terminal end to the 6XHis tag followed by the chitin binding/intein domain of the vector, designated pL7C. This and other subclone constructs are listed in the accompanying manuscript1.

Antibodies

Anti-*Gro*EL was purchased from Amersham-Pharmacia. Anti-intein binding domain was purchased from New England Biolabs.

Solubility Media

(w/v)-2% Bacto Tryptone, 0.8% NaCl, 1.5% Bacto Yeast Extract, 0.1%KH_2PO4, 0.2% $NaHPO_4.7H_2O$ has been reported to improve the solubility of some overexpressed phage T4 proteins27.

Chitin Column Sonication Buffer

0.4 mM $EDTA_{pH8}$; 20 mM $HEPES_{pH7.5}$ for GmrS or 20 mM $Tris.HCl_{pH8.0}$ for GMRd; 1.5 M NaCl for GMRs or 0.8 M NaCl for GmrD.

Chitin Column Equilibration Buffer

0.4 mM $EDTA_{pH8}$; 20 mM $HEPES_{pH7.5}$ for GmrS or 20 mM $Tris.HCl_{pH8.0}$ for GmrD; 0.1% Triton X-100; 1.5 M NaCl for purification of the GmrS protein or 0.8 M NaCl for purification of the GmrD protein.

Chitin Column Cleavage Buffer

0.4 mM $EDTA_{pH8}$; 20 mM $HEPES_{pH7.5}$ for GmrS or 20 mM $Tris.HCl_{pH8.0}$ for GmrD; 40 mM DTT (in $NaOAc_{pH5.5}$); 0.5 M NaCl for GmrS or 0.1 M NaCl for GmrD.

Chitin Column Elution Buffer

GMRs: 0.4 mM $EDTA_{pH8}$; 20 mM $HEPES_{pH7.5}$ for GmrS or 20 mM $Tris.HCl_{pH8.0}$ for GmrD; 0.1 M NaCl; 0.4 mM $EDTA_{pH8}$.

DNA Binding Buffer

Manufacturer (Roche) provided 5XBinding Buffer (20 mM $HEPES_{pH7.6}$, 1 mM EDTA, 10 mM $(NH_4)_2SO_4$, 1 mM DTT, 0.2% Tween-20, 30 mM KCl); 1 mM UTP; 3 mM $MgCl_2$; 5 mM $CaCl_2$.

Co-immunoprecipitation Buffer

0.1% TritonX-100; 100 mM NaCl; 5 mM $CaCl_2$; 3 mM $MgCl_2$; 2 mM KOAc; 20 mM $HEPES_{pH7.5}$; 1 mM UTP.

Methods

Phage isolation

Phage isolation and purification were performed as described in Sambrook *et.al.*198928

Glucosylation- and cytosine modification-deficient mutant phage production (amber α or β in combination with point mutations of glucosyl transferases)

Original stocks were prepared on sensitive, permissive *E. coli* strains as previously described. A non-amber suppressor strain, *sup*$^-$ (e.g. DH10B), and glucose donor-deficient hosts (e.g.

UDPG phosphorylase^{-} *E. coli* W4597 or the less fully *gal*U^{-} B834*gal*U56, and K803)26 were used for production of progeny phage with glc^{-} DNA. The non-amber suppressor DH10B was used for infection by T4GT7 for production of phage containing C DNA (non-modified cytosine)25. Glucosylation deficient phages were prepared as described29. Phage DNAs were extracted following purification of CsCl banded phage particles by standard techniques28. All DNAs were assayed by the ability of the following restriction endonucleases to digest: *Eco*RV and *Dra*I are able to digest glc-HMC T4 DNA, *Eco*RI is blocked by the glucosylation of HMC, *Hpa*II is blocked by methylation of cytosine, *Msp*I is blocked by methylation of cytosine, *Bgl*II is blocked by *dam* methylation and *Alu*I is blocked by the HMC modification30.

Intein/CBD Tag-based Purification of GmrS and GmrD

A number of purifications of the two proteins were attempted (His-Tag, GST-tag, Vpr-tag and ion exchange) and failed due to insolubility, inability to purify to an acceptable level, the lack of enzyme activity in the presence of a tag as well as susceptibility of the two proteins to proteolysis by the endoproteinase used to remove the purification tags. The only purification vehicle that met our needs was the chitin binding-based IMPACT-TWIN system from New England Biolabs. Overnight cultures of TSC (g*mrS* cloned into pTYB2) and TDC (*gmrD* cloned into pTYB2) colonies were diluted into 500 mL of glucose-free Solubility Media[34], grown for 1 h at 37 °C with vigorous shaking, moved to 29 °C for 20 min and induced with 0.5 mM IPTG. Growth was continued for 2 h 45 min after which the cells were harvested. TSC cells were re-suspended in 10 mL S-sonication buffer to which 1 mM PMSF, 1/50 volume protease inhibitor cocktail (Sigma), 10 mM $MgCl_2$, 20 nM GTP (Pharmacia) and 5 μg/mL protease-free Pancreatic DNaseI (Sigma) were added. The cell suspension was cooled on ice for 15 min prior to sonication at 32% power for a total 4 min 45 sec with 15sec intervals between sonication and rest using the Fisher Sonic Dismembrator Model 500. Immediately following sonication, TritonX-100 (Sigma) was added to a concentration of 0.1%, from a 10% stock. TDC cells were re-suspended in 10 mL D-Sonication buffer and processed as TSC with the exception that GTP was added at a concentration of 50 nM, no DNase was added and sonication was performed at a level of 34%. Cell rupture of both TSC and TDC suspensions were assayed using the hemocytometer, and by following the OD_{600} for the experimentally determined appropriate loss of culture density.

The ruptured cell suspensions were cooled on ice for 15 min prior to centrifugation in a Sorvall RC5B at 15,000 x g for 10 min. The supernatants were diluted by one volume with column buffer and loaded onto individual 0.8 mL chitin columns (NEB) equilibrated in either S- or D-Column buffer. The protein samples were loaded and 16 V of column buffer were used to wash off non-specifically binding proteins under gravity flow. One volume of S- or D-Cleavage buffer was loaded onto the two columns, and flow was stopped as the buffer was cleared. Complete DTT-catalyzed removal of the Intein/CBD tag unit required 16 h, and the partly purified proteins were eluted from the resin in either S-Elution Buffer or D-Elution buffer in 800 μL fractions. The GmrS protein eluted from the chitin column was dialyzed into S-Storage Buffer (20 mM $HEPES_{pH7.5}$, 500 mM NaCl, 1 mM EDTA, 1 mM DTT), and GmrD against D-Storage Buffer (20 mM Tris $HCl_{pH8.0}$, 130 mM NaCl, 1 mM EDTA, 1 mM DTT). Neither storage buffer contains glycerol as concentration of this standard reagent above 5% inactivates one or both of the proteins. The 500 mL preparations provided GmrS at up to 4.0 mg/mL and GmrD at 3.0 mg/mL, mixed with the chitin resin-binding *E. coli Gro*EL at ~200 mg/mL. This contamination by GroEL was also observed in the purification of other enzymes[35], though removal from either GmrS or GmrD could not be accomplished.

For studies of the GmrD-6XHis variant protein, the pL7C construct that contains the sequences *gmrD*-6XHis(*cac*)-intein/CBD in the pTYB2 expression vector, was used for GmrD-6XHis variant protein expression and purification. This double-tagged protein preparation utilized the

CBD tag for purification on a chitin resin column, and this protein eluted from the chitin column using DTT as described before.

GmrSD Nuclease assays

Immediately prior to use, proteins were spot-dialyzed on Millipore 0.025 μm dialysis disks against 30 mL of 25 mM $HEPES_{pH7.5}$ in a petri dish for 30 min at room temperature. Of note, incubating the two proteins on the dialysis disk together resulted in loss of activity and, as visualized on a 10% SDS polyacrylamide gel, breakdown of the GmrS protein. Each nuclease reaction contained 0.4 μg GmrS and/or 1.5 μg GmrD with approximately 70 ng DNA, 20 mM $Tris{\cdot}HCl_{pH7.5}$, 5 mM $CaCl_2$, 3 mM $MgCl_2$, 2 mM KOAc, 1 mM $UTP_{pH7.0}$ (MBI Fermentes) as the standard activation reagent. One millimolar GTP-γ-S $_{pH7.0}$ was used as the non-hydrolyzable control (MBI Fermentes). A Low Energy Environment nuclease reaction was developed that contains 1/2 the proteins and 1/2 the nucleotide used in the standard reaction, and the incubation time was extended to 5 h.

Size-exclusion chromatography

A size-exclusion column was used to assess the weight of enzymes. GmrS and GmrD were combined at the determined active 4:1 ratio and passed over FPLC Superose 6 resin pre-packed in an HR 10/30 column at a speed of 0.5 mL/min, using the BioRad Biologic Chromatography System. The column and proteins were equilibrated in buffer containing 100 mM NaCl, 30 mM $MgCl_2$, 20 mM KOAc and 20 mM $HEPES_{pH7.5}$. The eluants were monitored spectrographically, and all peak fractions were TCA-precipitated and run on a 10% SDS polyacrylamide gel as before.

Co-immunoprecipitation

Seven micrograms of dialyzed GmrS and/or 7 μg dialyzed GmrD, 5 μg IPI or 10^9 burst phages and 1:2,000 IPI antiserum γ–globulin antibody were combined in co-immunoprecipitation buffer +/− 1 mM UTP. IPI* antiserum was prepared against carboxymethylcellulose column chromatographed IPI* from purified wild type phage T4. Standard protocols for co-immunoprecpitation were followed.

Affinity chromatography of GmrS and GmrS

Further investigation of the potential of physical association between the IPI* protein with either GmrS or GmrD was performed utilizing a His-tag IPI* fusion protein immobilized on a metal-affinity column (in collaboration with Dr. Rifat). The column was equilibrated with 50 mM $KHPO_{4\ pH7.4,\ 0.1}$ M NaCl, 0.02 mM imidazole, 3 mM $MgCl_2$, 2 mM KOAc; and the proteins were dialyzed into this same buffer. Prior to applying the proteins to the column, T4 DNA, 1 mM UTP and 5 mM $CaCl_2$ were combined with the three proteins and incubated at 4 °C for 10 min. The GmrS and GmrD proteins were incubated in equimolar amounts with 6XHis-IPI* and the reaction components were separated over nickel-charged resin. The resin was washed with 10 V of the equilibration buffer, and serial elution was accomplished with increasing imidazole concentrations (0.02 M, 0.09 M, 0.16 M, 0.25 M imidazole).

References

1. Bair, CL.; Rifat, D.; Black, LW. accompanying manuscript. 2006. Exclusion of glucosyl-hydroxymethylcytosine DNA containing bacteriophages is overcome by the injected protein inhibitor IPI*.
2. Bickle, TA. DNA restriction and modification systems. In E. coli and S. typhimurium. In: Neidhardt, H., editor. Cellular and Molecular Biology. ASM Press; Washington DC: 1987. p. 692-696.
3. Roberts RJ, Belfort M, et al. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucl Acids Res 2003;31:1805–12. [PubMed: 12654995]

4. Modrich, P.; Roberts, RJ. Type II RMS. In: Lin, SM.; Roberts, RJ., editors. Nucleases. Cold spring Harbor Lab Press; Cold Spring Harbor, NY: 1982. p. 109-154.
5. Murray NE. Type I restriction systems: sophisticated molecular machines (a Legacy of Bertani and Weigle). Microbiol Mol Bio Rev 2000;64:412–434. [PubMed: 10839821]
6. Bickle TA, Krüger DH. The biology of DNA restriction. Microbiol Mol Bio Rev 1993;57:434–450.
7. Roberts, RJ.; Macelis, D. The restriction enzymes. In: Linn, SM.; Roberts, RJ., editors. Nucleases. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, NY: 1993. p. 439-444.
8. Sutherland E, Coe L, Raleigh EA. McrBC: a multisubunit GTP-dependent resrtiction endonuclease. J Mol Biol 1992;225:327–348. [PubMed: 1317461]
9. Lepikhov K, Tchernov A, Zheleznaja L, Matvienko N, Walter J, Trautner TA. Characterization of the Type IV restriction modification system BspLU11III from *Bacillus* sp LU11. Nucl Acids Res 2001;29:4691–4698. [PubMed: 11713319]
10. Carlson, K.; Raleigh, EA.; Hattman, S. Restriction and Modification. In: Karam, JD., editor. Molecular Biology of bacteriophage T4. American Society for Molecular Biology; Washington, D.C: 1994. p. 369-381.
11. Raleigh EA, Brenner J, Bloom F, Braymer HD, DeCruz E, Dharmalingam K, Heitman J, Noyer WM, Piekarowicz A, Kretz PL, Short JM, Woodcock D. Nomenclature relating to restriction of modified DNA in *E. coli.* J Bacteriol 1991;173:2707–09. [PubMed: 2013582]
12. Weinstock GM, McEntee K, Lehman IR. Hydrolysis of nucleoside triphosphates catalyzed by the recA protein of *Escherichia coli* Characterization of ATP hydrolysis. J Biol Chem 1981;256:8829–8834. [PubMed: 7021552]
13. Kuno S, Lehman IR. Gentiobiose, a constituent of deoxyribonucleic acid from coliphage T6. J Biol Chem 1962;237:1266–1270. [PubMed: 14460593]
14. Lehman IR, Pratt EA. On the structure of the glucosylated hydroxymethylcytosine nucleotides of coliphages T2, T4, and T6. J Biol Chem 1960;235:3254–3259. [PubMed: 13760441]
15. Georgopoulos CP, Revel HR. Studies of glucosyltransferase mutants of the Teven bacteriophages. Virology 1971;44:271–85. [PubMed: 4935677]
16. Black, LW.; Showe, MK.; Steven, AC. Morphogenesis of the T4 Head. In: Karam, JD., editor. Molecular Biology of bacteriophage T4. American Society for Molecular Biology; Washington, D.C: 1994. p. 369-381.
17. Bourniquel AA, Bickle TA. Complex restriction enzymes: NTP-driven molecular motors. Biochimie 2002;84:1047–59. [PubMed: 12595133]
18. Panne D, Raleigh EA, Bickle T. McrBS, a modulator peptide for McrBC activity. EMBO J 1998;17:5477–5483. [PubMed: 9736625]
19. Panne D, Müller SA, Wirtz A, Engel A, Bickle TA. The McrBC restriction endonuclease assembles into a ring structure in the presence of G nucleotides. EMBO J 2001;20:3210–3217. [PubMed: 11406597]
20. Stewart FJ, Panne D, Bickle TA, Raleigh EA. Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. J Mol Biol 2000;298:611–22. [PubMed: 10788324]
21. Ellis DJ, Dryden DT, Berge T, Edwardson JM, Henderson RM. Direct observation of DNA translocation and cleavage by the *Eco*KI endonuclease using atomic force microscopy. Nat Struct Biol 1999;6:15–17. [PubMed: 9886284]
22. Dila D, Sutherland E, Moran L, Slatko B, Raleigh EA. Genetic and sequence organization of the g*mcrbc* locus of *Escherichia coli* K-12. J Bacteriol 1990;172:4888–4900. [PubMed: 2203735]
23. Pieper U, Groll DH, Wunsch S, Gast FU, Speck C, Mucke N, Pingoud A. The GTP-dependent restriction enzyme McrBC from *Escherichia coli* forms high-molecular mass complexes with DNA and produces a cleavage pattern with a characteristic 10-base pair repeat. Biochem 2002;41:5245–54. [PubMed: 11955074]
24. Wilson GG. Organization of restriction-modification systems. Nucl Acids Res 1991;19:2539–2566. [PubMed: 2041731]
25. Wilson GG, Young KKG, Edlin GJ, Konigsberg W. High-frequency generalized transduction by bacteriophage T4. Nature (London) 1979;280:80–81. [PubMed: 15305587]

26. Runnels J, Snyder L. Isolation of a bacterial host selective for bacteriophage T4 containing cytosine in its DNA. J Virol 1978;27:815–8. [PubMed: 359832]

27. Moore JT, Uppal A, Maley F, Maley GF. Overcoming Inclusion Body Formation in a High-Level Expression System. Prot Express Purif 1993;4:160–3.

28. Sambrook, J.; Maniatis, T.; Fritsch, EF. Molecular Cloning: A Laboratory Manual. 2. Sambrook, J., editor. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, NY: 1989.

29. Hattman S, Fukasawa T. Host-induced modification of Teven phages due to defective glucosylation of their DNA. Proc Natl Acad Sci (USA) 1963;50:287–300.

30. Kutter, E.; Stidman, T.; Guttman, B.; Kutter, E.; Batts, D.; Peterson, S.; Djavakhishvili, T.; Arisaka, F.; Mesyanzhinov, V.; Rueger, W.; Mosig, G. Genomic Map of Bacteriohage T4. In: Karam, JD., editor. Molecular Biology of bacteriophage T4. American Society for Molecular Biology; Washington, D.C: 1994. p. 369-381.

31. Janosi L, Yonemitsu H, Hong H, Kaji A. Molecular cloning and expression of a novel hydroxymethylcytosine-specific restriction enzyme (*Pvu*Rts1I) modulated by glucosylation of DNA. J Mol Biol 1994;242:45–61. [PubMed: 8078071]

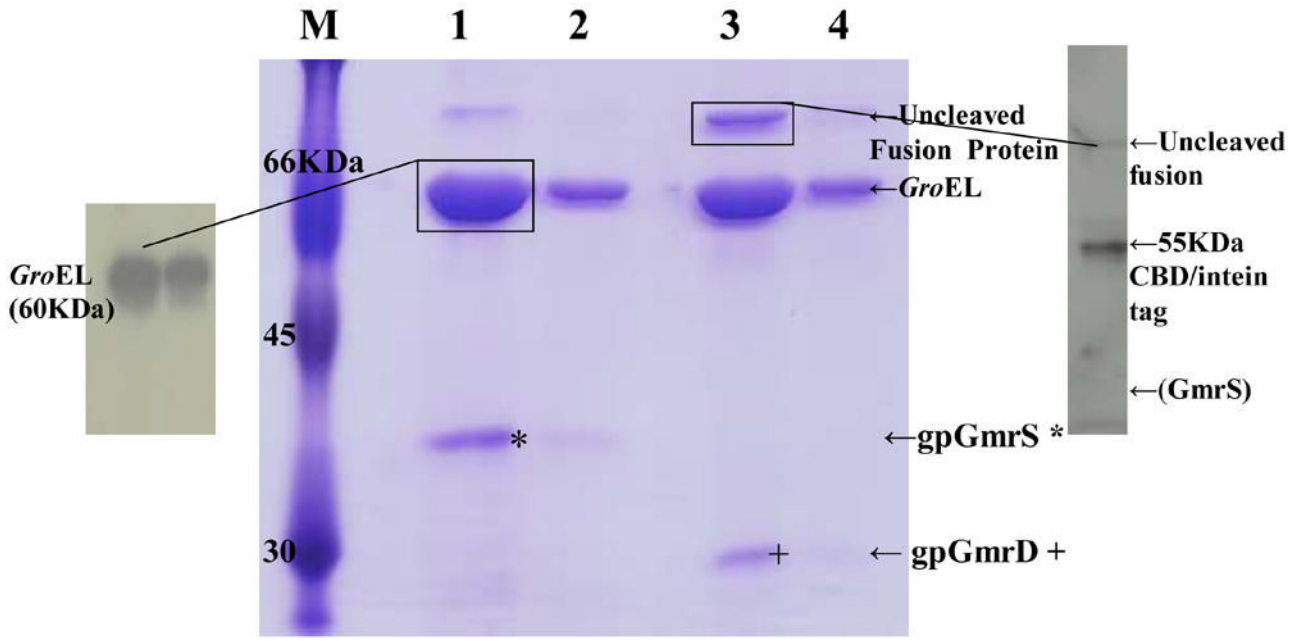

Figure 1. Purification by affinity chromatography of GmrS and GmrD following expression of the chitin binding domain (CBD) intein self-cleaving fusion proteins

M) Protein Markers; 1) pTSC (Impact-CN *gmrS* expressed GmrS) Eluate I; 2) pTSC Eluate II; 3) pTDC (Impact-CN *gmrD* expressed GmrD) Eluate I; 4) pTDC Eluate II. The major contaminants *Gro*EL and the uncleaved CBD/intein fusion protein are identified by Western blot, and the purified proteins are run on an SDS-PAGE stained with Coomassie blue.

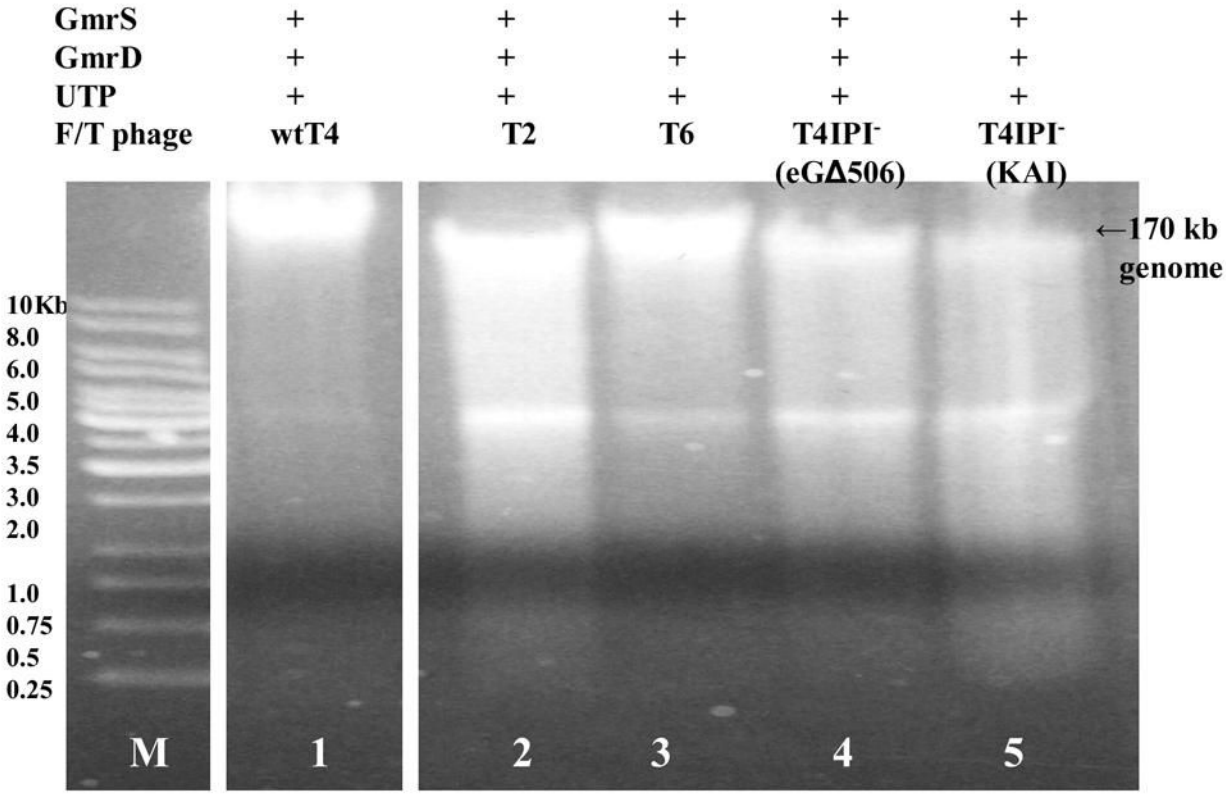

Figure 2. The GmrSD enzyme digests the DNAs of phages lacking IPI*

The DNAs and internal proteins from pure phage particles were released by freeze/thawing, subjected to GmrSD digestion in the presence of 0.5 mM UTP, and run on an agarose gel stained with EthBr. M) DNA markers; 1) wtT4; 2) T2; 3) T6; 4) T4 *ip1*$^{-}$ deletion mutant eG506; 5) T4 *ip1*$^{-}$ point mutant (KAI^{-}). The DNAs are run on an agarose gel stained with EthBr.

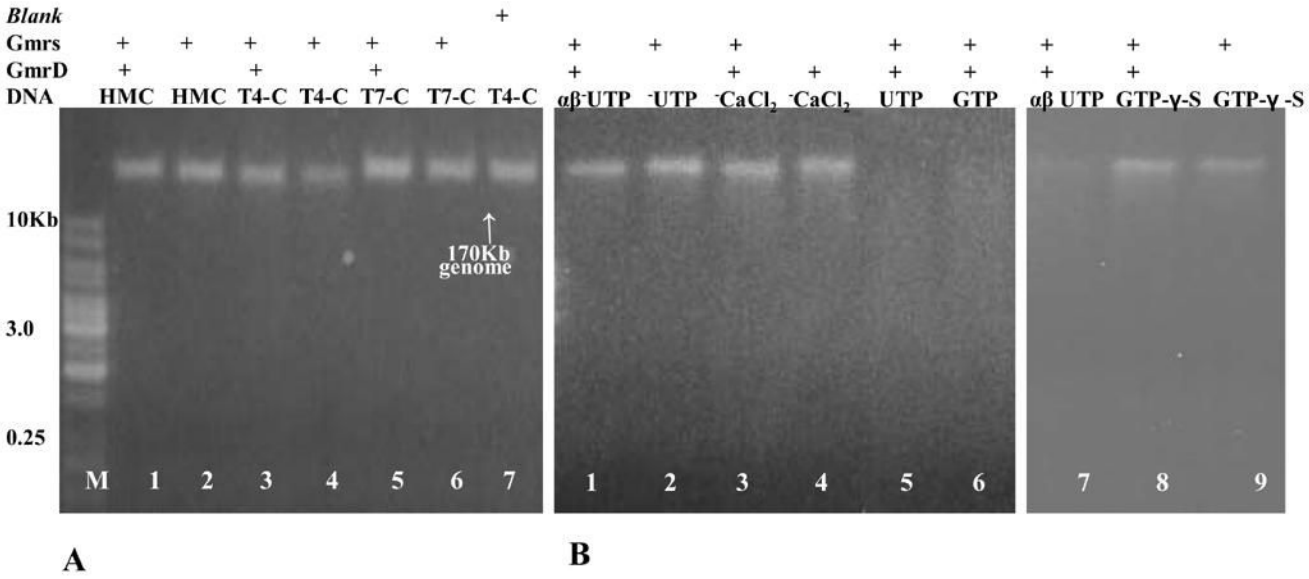

Figure 3. The GmrSD enzyme does not digest purified phage DNAs lacking modified cytosines and requires calcium and NTP hydrolysis for activity

Panel A: Lane 1) GmrS + T4 HMC DNA; 2) GmrSD + T4 HMC DNA; 3) GmrS + T4 C DNA; 4) GmrSD + T4 C DNA; 5) GmrS + T7 DNA; 6) GmrSD + T7 DNA; 7) *Gro*EL Blank + T4 C DNA. Panel B: 1) GmrSD + T4 DNA (−UTP); 2) GmrS + T4DNA (−UTP); 3) GmrSD + T4 DNA (−$CaCl_2$); 4) GmrD + T4 DNA (−$CaCl_2$); 5) GmrSD + α-glc HMC DNA; 6) GmrSD + α-glc HMC DNA + GTP; 7) GmrSD + T4 DNA + UTP; 8) GmrSD + T4 DNA +GTP-γ-S; 9) GmrS + T4 DNA + GTP-γ-S. All reactions were performed at 37 °C for 30 min in 1 mM UTP, 2 mM Kacetate, 3 mM $MgCl_2$ and 5 mM $CaCl_2$ except where indicated. The DNAs are run on an agarose gel and stained with EthBr.

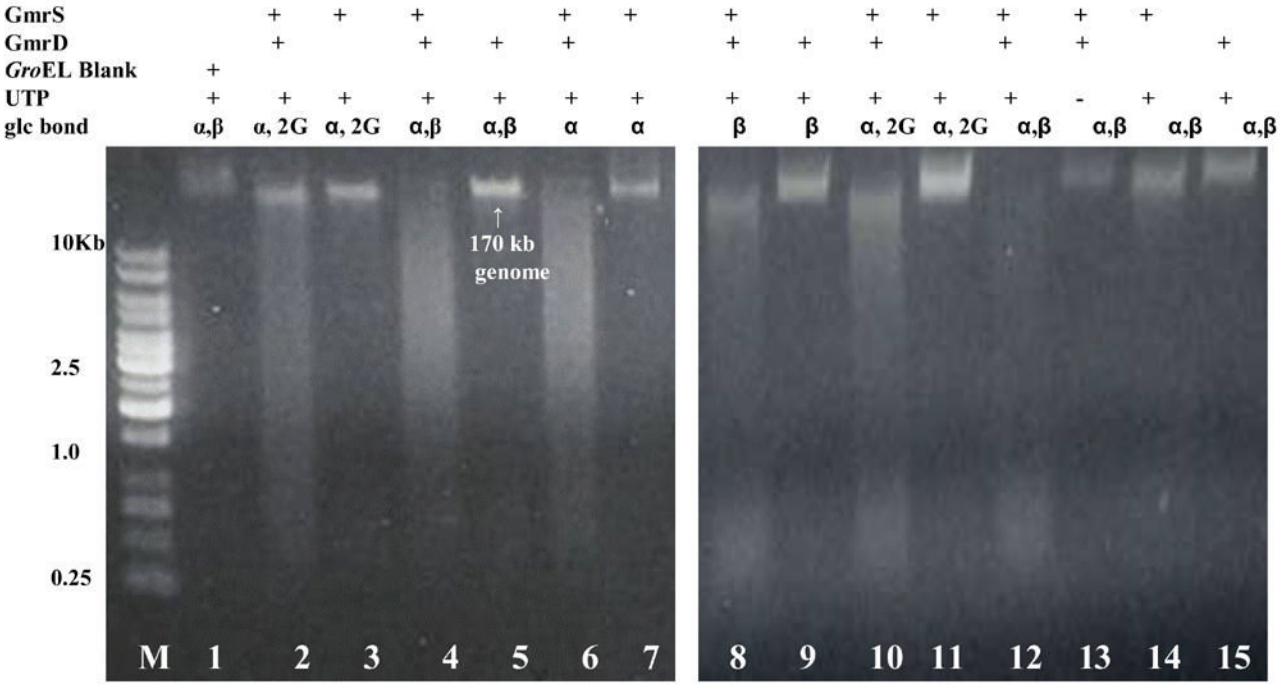

Figure 4. The GmrSD enzyme is able to digest α-glc, β-glc, and gentiobiosyl modified HMC containing T-even phage DNAs

M) DNA markers; 1) *Gro*EL Blank + wtT4 DNA; 2) GmrSD + T2 DNA; 3) GmrS + T2 DNA; 4) GmrSD + T4 DNA; 5) GmrD + wtT4 DNA; 6) GmrSD + α-glcT4 DNA; 7) GmrS + α-glcT4 DNA; 8) GmrSD + β-glcT4 DNA; 9) GmrD + β-glcT4 DNA; 10) GmrSD + T6 DNA; 11) GmrS + T6 DNA; 12) GmrSD + wtT4 DNA; 13) GmrSD + wtT4 DNA(−UTP); 14) GmrS + wtT4 DNA; 15) GmrD + wtT4 DNA. All reactions were carried out at 37 °C for 30 min in 1 mM UTP, 2 mM Kacetate, 3 mM $MgCl_2$ and 5 mM $CaCl_2$ except where indicated. The DNAs are run on an agarose gel and stained with EthBr.

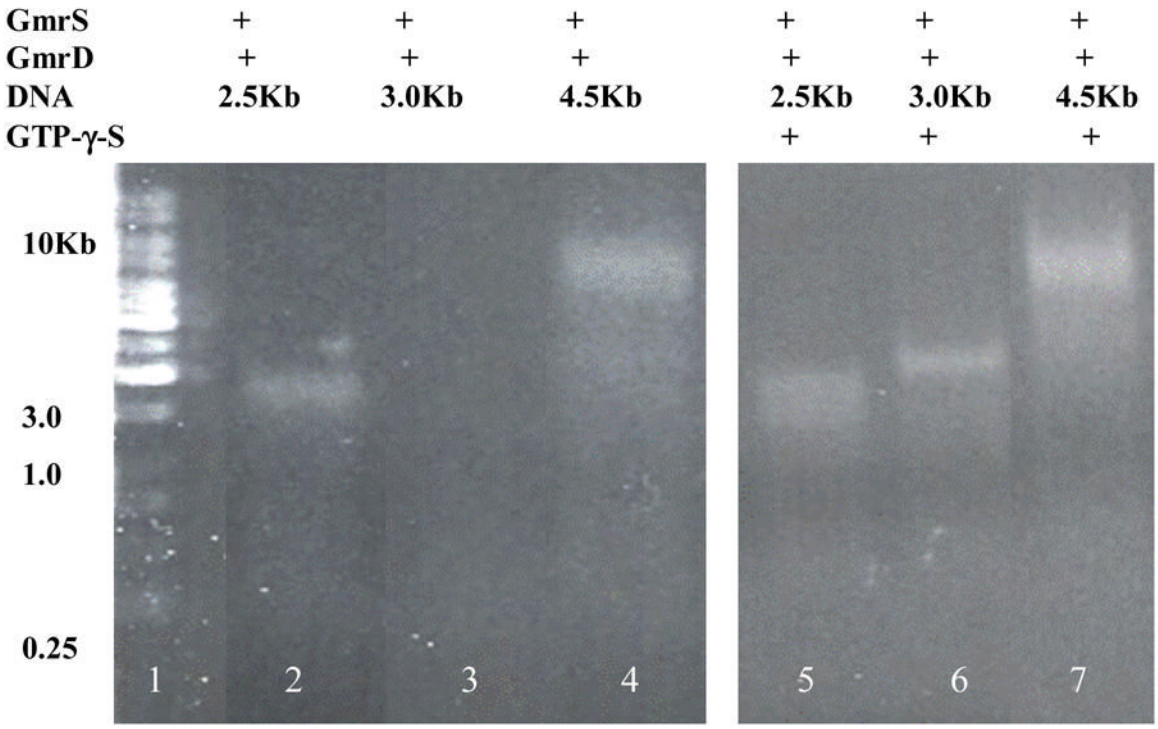

Figure 5. The GmrSD enzyme digests the 3 kb but not the 4.5 kb and 2.5 kb wtT4 DNA *Taq*I fragments

All reactions contain GmrS and GmrD and 1 mM UTP. M) DNA markers; 1) 2.5 kb fragment; 2) 3 kb fragment; 3) 4.5 kb fragment; 4) 2.5 kb + GTP-Y-S; 5) 3.0 kb + GTP-Y-S; 6) 4.5 kb + GTP-Y-S. All reactions were performed at 37 °C for 30 min in 1 mM UTP, 2 mM Kacetate, 3 mM $MgCl_2$ and 5 mM $CaCl_2$, or the inhibitory analogue GTP-γ-S where noted. The DNAs are run on an agarose gel and stained with EthBr.

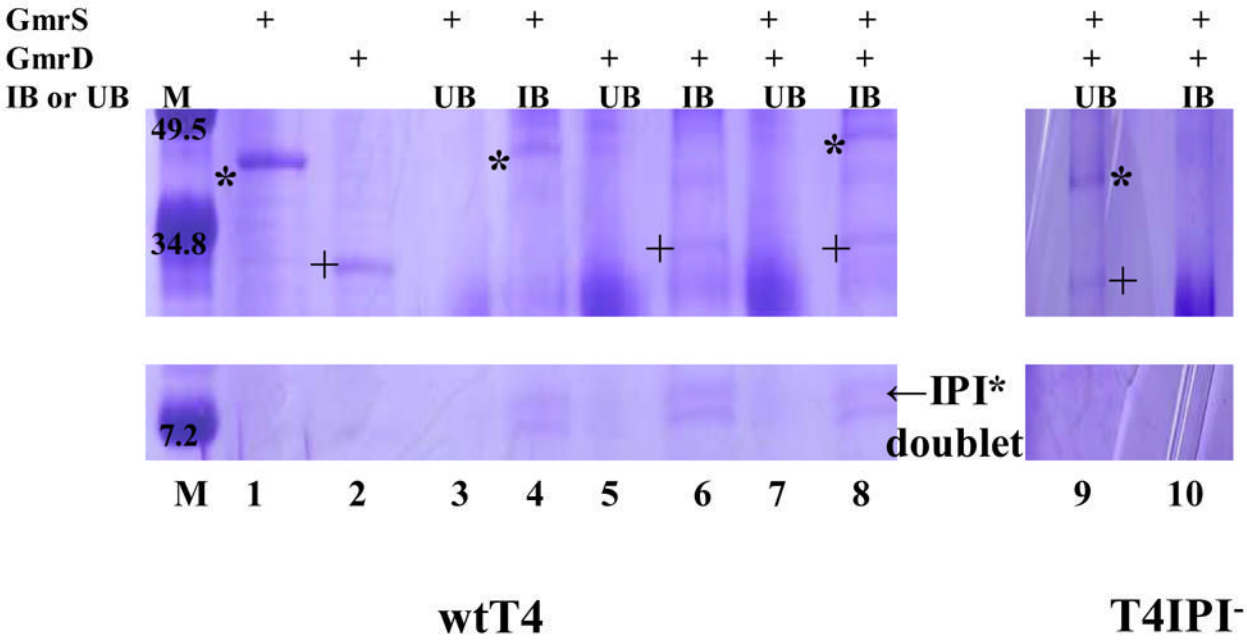

Figure 6. The GmrS and GmrD proteins are both co-immunoprecipitated with the IPI* protein released from frozen/thawed wt T4 phage particles (wt T4 left panels) but not disrupted T4 *ip1*- phage lacking IPI* (8 KDa) (T4IPI^{-} right panels)

Anti-IPI* IgG is used for the immunoprecipitation. Lanes are labeled either IPI Bound (IB) for proteins immunoprecipitated or IPI Unbound (UB) for those proteins not immunoprecipitated through an association with IPI*. M) protein markers; 1) GmrS* input protein; 2) GmrD + input protein; 3) Unbound + GmrS; 4) Bound + GmrS; 5) Unbound + GmrD; 6) Bound + GmrD; 7) Unbound + GmrSD; 8) Bound + GmrSD; 9) Unbound –IPI* + GmrSD; 10) Bound –IPI* + GmrSD. The immunoprecipitates and supernatants are run on SDS-PAGE and stained with coomassie blue.

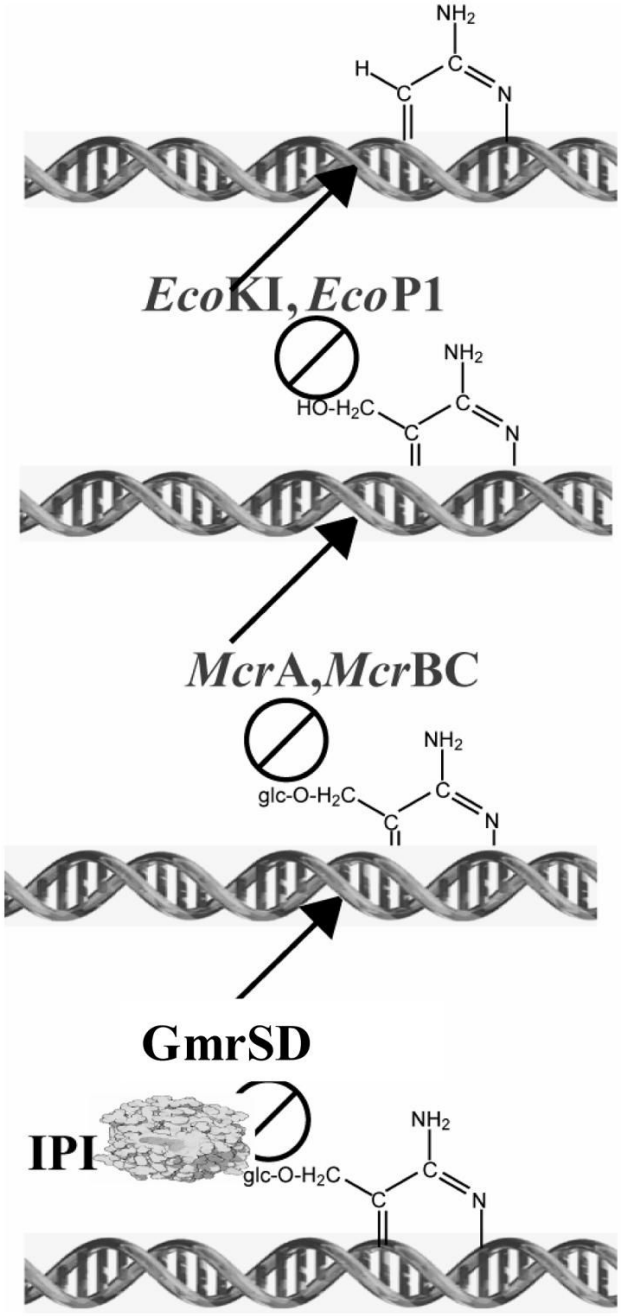

Fig. 7. Evolution of *Myoviridae* DNA Modifications and of GmrSD and other DNA Modification Dependent Restriction *Endonucleases*

Restriction endonucleases encoded by *E. coli* K12, Prophage P1 and numerous bacteria protect against infecting phage DNAs containing cytosine (first line). Many such enzymes are blocked by methylation or hydroxymethylation (HMC) of cytosine (line 2). The *Mcr*A and *Mcr*BC modification dependent restriction endonucleases of *E. coli* specifically attack HMC modified *Myoviridae* DNA, but are inhibited by the glucosylation of HMC (glc-HMC) (line 3). The GmrSD enzyme is able to digest the sugar modified (glc)-HMC containing DNAs of a number of T-even phages, but its activity is inhibited by the encapsidated phage IPI* protein injected with the DNA (line 4). Polymorphism of the *ip1* gene family and of HMC sugar modifications

among the *Myoviridae* suggest a widespread and incompletely characterized *Gmr* enzyme family.