PrfA protein of *Bacillus* species: Prediction and demonstration of endonuclease activity on DNA

DANIEL J. RIGDEN,¹ PETER SETLOW,² BARBARA SETLOW,² IRINA BAGYAN,²
RICHARD A. STEIN,³ AND MARK J. JEDRZEJAS^{3,4}

¹National Centre of Genetic Resources and Biotechnology, Cenargen/Embrapa, Brasília, Brazil, D.F. 70770–900

²Department of Biochemistry, University of Connecticut Health Center, Farmington, Connecticut 06032, USA

³Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA

⁴Children's Hospital Oakland Research Institute, Oakland, California 94609, USA

(RECEIVED May 22, 2002; FINAL REVISION July 9, 2002; ACCEPTED July 9, 2002)

Abstract

The *prfA* gene product of Gram-positive bacteria is unusual in being implicated in several cellular processes; cell wall synthesis, chromosome segregation, and DNA recombination and repair. However, no homology of PrfA with other proteins has been evident. Here we report a structural relationship between PrfA and the restriction enzyme *PvuII*, and thereby produce models that predict that PrfA binds DNA. Indeed, wild-type *Bacillus stearothermophilus* PrfA, but not a catalytic site mutant, nicked one strand of supercoiled plasmid templates leaving 5'-phosphate and 3'-hydroxyl termini. This activity, much lower on linear or relaxed circular double-stranded DNA or on single-stranded DNA, is consistent with a role for this protein in chromosome segregation, DNA recombination, or DNA repair.

Keywords: Endonuclease activity; fold recognition; function prediction; homology modeling

The *prfA* gene was originally identified through its cotranscription with the *Bacillus subtilis* *ponA* gene (Popham and Setlow 1995). The latter gene encodes a member of the Class A penicillin-binding protein family that possesses transglycosylase and transpeptidase activities (Ghuysen 1994). The presence of the *prfA* gene in an operon with *ponA* suggested a role for PrfA in cell wall synthesis. Indeed, inactivation of *prfA* reduces the rate of cell growth around 50%, while simultaneous mutation of *ponA* and *prfA* has much more dramatic effects on cell growth and sporulation than either individual mutation (Popham and Setlow 1995). The location of *prfA* and *ponA* in an operon in *Staphylococcus aureus* further reinforced the idea of a functional link between the products of these two genes (Pinho et al. 1998). It was therefore surprising when mutational data showed that *prfA* (under its alternative name of *recU*) was

required for DNA repair and recombination in *B. subtilis* (Fernandez et al. 1998). PrfA was also shown to be necessary for proper chromosome segregation in *B. subtilis* (Pedersen and Setlow 2000). This latter finding is consistent with a role for PrfA in recombination, because mutations in other genes involved in homologous recombination also lead to defects in chromosome segregation (Wake and Errington 1995).

Further study of the PrfA protein and its precise role in vivo has been severely hampered by the lack of significant sequence identity between PrfA and proteins of known structure or function. Here we report fold recognition and sequence analysis results that support a structural correspondence between PrfA and *PvuII*, a type II restriction enzyme. This class of enzymes shares structural similarities in the absence of significant sequence similarity (Kovall and Matthews 1999). Recent sequence (Aravind et al. 2000) and structural (e.g., Ban and Yang 1998) analyses place type II restriction enzymes in an extended superfamily containing DNA and RNA endonucleases and exonucleases involved in diverse cellular functions, among them several steps of DNA recombination and repair. Although the original threading alignment of PrfA led to poor quality models, application of rigorously tested alignment shifts produced

Reprint requests to: Daniel J. Rigden, National Centre of Genetic Resources and Biotechnology, Cenargen/Embrapa, Brasília, Brazil, D.F. 70770-900; e-mail: daniel@cenargen.embrapa.br; fax: 55 (61)340-3658 or Mark J. Jedrzejewski, Children's Hospital Oakland Research Institute, 5700 Martin Luther King Jr. Way, Oakland, CA 94609, USA; e-mail: mjedrzejewski@chori.org; fax: (510) 450-7910.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.0216802>.

dramatic improvements in model reliability. PrfA is therefore confidently identified as a new member of the type II restriction enzyme superfamily.

The structural correspondence between PrfA and members of the type II restriction enzyme superfamily strongly suggested that PrfA would have some enzymatic activity on DNA. Although previous work (Pedersen and Setlow 2000) has shown that PrfA has no DNA helicase activity, we now demonstrate that PrfA does have endonuclease activity on DNA, and makes single strand breaks with 5'-phosphate and 3'-hydroxyl termini in supercoiled plasmid DNA, but has much lower activity on relaxed double-stranded plasmid or linear DNA or single-stranded DNA. The endonuclease activity of PrfA on DNA is consistent with a role for this protein in chromosome segregation or DNA recombination or repair.

Results and Discussion

Sequence relationships

The principal aim of this work was the construction of a three-dimensional (3D) model of the PrfA protein of *Bacillus*

subtilis based on the analysis of PrfA of *B. subtilis* and *B. stearothermophilus*. The amino acid sequence relationships between known PrfA homologs are illustrated in Figures 1–2. The PrfA homologs are a diverse sequence group, sharing a mean pairwise sequence identity of just 36%. Although PrfA was identified first from its putative involvement in cell wall synthesis (Popham and Setlow 1995), PrfA homologs are present in organisms lacking cell walls, including *Mycobacterium genitalium* and *Ureaplasma urealyticum*. This result strongly suggests that the principal *in vivo* role of PrfA may be during recombination (Fernandez et al. 1998). The discovery of a sequence homologous to PrfA in the *B. subtilis* phage SPBc2 was somewhat surprising, although bacteriophage do encode enzymes involved in recombination (Lilley and White 2001; Sharples 2001). However, this phage PrfA homolog may well lack activity and have arisen by chance through gene duplication. The evidence for this is severalfold: (1) alignment, pairwise comparisons, and MEME analysis all show this phage sequence to be the most atypical PrfA homolog, having, for example, many insertions and deletions not present in other sequences (Fig. 1); (2) the pairwise sequence identity of the

Fig. 1. Alignment of PrfA sequences and homologs with *Pvu* II and *SptAIR*. Shaded regions indicate identities within either the PrfA sequences or between the two restriction enzymes. Emboldened positions are those conserved in at least 12 of the 13 sequences. The *B. subtilis* and *B. stearothermophilus* PrfA sequences and the *Pvu*II sequence are numbered above and below the alignment, respectively. Predicted secondary structures for *B. subtilis* PrfA are shown at the top of the sequences and the actual secondary structure of leyu at the bottom. MEME motifs are marked and numbered in approximate order of significance. The figure was produced using ALSCRIPT (Barton 1993).

Fig. 2. Neighbor-joining tree representation of the relationships between PrfAs and homologous sequences produced with the PHYLIP package (Felsenstein 1989).

phage sequence with the other PrfAs is just 14–22% (Fig. 1); and (3) involvement in the phage life cycle appears unlikely, because none of the other 21 phage genome sequences with the same taxonomy (tailed phages: siphoviridae—obtained at <http://www.ncbi.nlm.nih.gov/ICTV/>) contain a PrfA homolog. Curiously, the genomic sequence of *Bacillus halodurans* encodes two PrfA homologs, one annotated as *prfA* (number 2 in Figs. 1–2) and the other unannotated (number 1). Sequence analysis clearly suggests that the unannotated *prfA* homolog is the more typical, and this gene also shares the positioning upstream of *ponA* with those of the single *prfA* homologs in *B. subtilis* (Popham and Setlow 1995), *B. stearothermophilus*, and *S. aureus* (Pinho et al. 1998). These sequence and gene location considerations, combined with the observation that no other bacterium yet sequenced encodes two PrfA homologs, suggest that homolog number 1 probably encodes the active PrfA of *B. halodurans* and that homolog number 2 may be an inactive consequence of gene duplication.

Hydrodynamic characterization

Band and boundary sedimentation velocity analyses were performed on the *B. subtilis* protein to investigate the properties of PrfA in solution and determine the correct oligomeric state for model construction. The sedimentation coefficient was first determined using band sedimentation

analysis with protein at 1.0 mg/mL and at two NaCl concentrations, 50 and 100 mM, which provided $s_{20,w}$ values of 2.95 and 2.96 S, respectively, corresponding to a dimeric species of PrfA (Table 1).

The distribution of $g(s^*)$ sedimentation coefficients from the boundary experiments also showed the presence of only dimeric PrfA with the averaged maximum peak height of $g(s^*)$ at $s_{20,w}$ of 3.06 S. (Table 1). The band and boundary sedimentation results allowed calculation of an average diffusion coefficient of $5.56 \cdot 10^{-7} \text{ cm}^2/\text{sec}$ and a molecular weight of ~50 kD for PrfA. These numbers are within the experimental errors when compared to *B. subtilis* PrfA's molecular weight of 24 kD per monomer (48 kD per dimer), calculated from its amino acid sequence. Both band and boundary sedimentation results are thus consistent with the presence of only a PrfA dimer in solution, as no traces of a monomer or higher molecular weight species were observed.

Fold recognition studies

Purely sequence-based methods, PSI-BLAST (Altschul et al. 1997) and HMM (Karplus et al. 1998), failed to find suitable templates for construction of a PrfA model. Consequently, the *B. subtilis* PrfA sequence was analyzed using a variety of Web-base threading servers. These methods employ a variety of inferred characteristics, such as predicted secondary structure and predicted solvent exposure to match sequences to known protein structures even in the absence of significant sequence similarity between them. The best scoring results were obtained using the 3D-PSSM

Fig. 3. Results of protein structure validation tools at different stages of *B. subtilis* PrfA model construction. (a–c) PROSA II, pG, and VERIFY_3D scores, respectively. Continuous lines indicate mean scores among sets of models produced using a given alignment and dashed lines the best score at that stage. The dotted line in (b) marks the threshold above which models are considered reliable in terms of template choice and alignment quality.

Table 1. Velocity band and boundary analytical sedimentation parameters for *B. subtilis* PrfA

Sedimentation velocity method	NaCl concentration (mM)	Corrected sedimentation coefficient, $s_{20,w}$ (Svedberg units) ^a	Diffusion coefficient, D (10^{-2} cm ² /s)	Determined molecular weight (kD)
Band	50	2.95	5.56	50,555
	100	2.96	5.57	50,779
Boundary, $g(s^*)$ analysis ^b	150	3.06	5.56	50,176
Average of the band and boundary analyses	Not applicable	2.99	5.56	50,503

The experiments were performed using 1.0 mg/ml PrfA at pH 6.0 as described in Experimental Procedures.

^a The partial specific volume, v , was calculated to be 0.730 based on the *B. subtilis* PrfA sequence.

^b The $g(s^*)$ was evaluated for several distribution data sets and then averaged.

method (Kelley et al. 2000) where both 1ygh (the structure of a non-DNA binding domain of transcriptional activator GCN5) and 3pvi (*PvuII* restriction enzyme) produced scores of -0.172 and 0.295 , respectively, below the “interesting” threshold of 1.0 . Given the phenotypically defined role of PrfA in recombination (Fernandez et al. 1998; Pedersen and Setlow 2000), the good-scoring putative structural correspondence with *PvuII* was suggestive. However, the alignment of PrfA with 3pvi did not cover the whole of the PrfA sequence. Because threading methods work best when supplied solely with the sequence matching the database structure, the threading experiments were repeated with the shorter putative match, lacking 25 residues not aligned with *PvuII*. By the 3D-PSSM method, 1pvi was now the only hit below the 1.0 threshold, albeit with a worse score of 0.457 and 1ygh was not present in the top 20 hits. Using this shorter PrfA sequence, structures for *PvuII* were also present in the top hits produced by other methods, having previously been absent. By the FFAS method (Rychlewski et al. 2000), 1eyu (*PvuII* endonuclease/cognate DNA complex) was the top-scoring hit at 5.23 , while 1pvi was 5th by the Bioinbgu consensus ranking (Fischer 2000) and top scoring by some of the individual methods.

To improve confidence in the putative structural match with *PvuII*, different PrfA homolog sequences were sent to the threading servers. Those chosen were the sequences from *Streptococcus pneumoniae*, the second *B. halodurans* homolog and the *U. urealyticum* sequence. These sequences share 46, 32, and 27% identity, respectively, with the *B. subtilis* sequence and at most 30% sequence identity among themselves, hence representing a highly diverse set. The *S. pneumoniae* sequence was again truncated to the size of the putative structural match, while the other two sequences match closely the size of *PvuII* (Fig. 1). With these sequences 3pvi was the top hit from 3D-PSSM analysis twice and was placed fourth on the third occasion. For the *U. urealyticum* sequence, FFAS and Bioinbgu consensus methods both also placed *PvuII* structures first.

Although the repeated structural match of PrfAs with *PvuII* suggested by independent fold recognition programs

for several diverse PrfA sequences was already highly suggestive, several further analyses increased confidence in the result significantly. First, although the majority of PrfA homologs are 20–25 residues longer than *PvuII* at the N-terminus, four sequences lack this extension and match closely the length of *PvuII*. Second, an excellent match is seen between the secondary structure predicted for *B. subtilis* PrfA by three independent methods, and the actual secondary structure of *PvuII* (Fig. 1). Predicted PrfA secondary structural elements not matched to the actual *PvuII* secondary structure are largely confined to regions in PrfA not present in *PvuII*—the N-terminal tail and two large insertions.

Strong support for a structural match between PrfA and *PvuII* also came from MEME analysis (Grundy et al. 1996). This method locates motifs—regions of particularly high sequence conservation—in sets of unaligned sequences. These motifs, maintained during evolution, may be expected to have structural or functional significance. The most highly conserved MEME motif present in the diverse set of PrfA homologs aligns with a region containing the catalytic triad of *PvuII*. These three *PvuII* residues, Asp 58, Glu 68, and Lys 70, are found in most members of the extended family of type II restriction endonucleases and related enzymes, and have been repeatedly highlighted by structural comparisons and sensitive sequence analyses (Kovall and Matthews 1999; Aravind et al. 2000; Declais et al. 2001). The first two residues are invariably conserved as acidic residues, and are involved in divalent cation binding. In contrast, the third residue, although generally a Lys, is more variable, with Glu and Gln present in the type II restriction enzymes *BamHI* and *BglII*, respectively. In the set of PrfA sequences, discounting the more distantly related *B. halodurans* homolog and the phage protein, both of doubtful activity (see above), these three positions are occupied by completely conserved Asp, Asp, and Glu residues. The second ranked MEME motif covers the long N-terminal helix of *PvuII*, which forms most of its dimer interface while motifs 3–4 match regions of *PvuII* containing DNA contacts (Fig. 1).

In an alignment containing all the PrfA homologs, *PvuII* and the only identifiable *PvuII* homolog, the SptAIR endonuclease from *Salmonella paratyphi* (Genbank accession AAG42427), very few residues were highly conserved (Fig. 1). Nevertheless, putative structural or functional roles could be assigned to these residues. As well as Asp 88 (*B. subtilis* PrfA numbering) that was part of the putative catalytic triad (see above), Gly 95, His 119, and Phe 135 were conserved, or only varied in the phage sequence. Corresponding residues in the *PvuII* structure are interesting because they are located in a disallowed area of the Ramachandran plot, are involved in key dimer–interface interaction (Cheng et al. 1994), and form the heart of the hydrophobic core of the catalytic domain, respectively. The special roles that can be assigned to the few highly conserved residues are suggestive of a distant evolutionary relationship and hence support the putative PrfA-*PvuII* structural correspondence.

The final piece of support for the PrfA-*PvuII* structural match came from consideration of the activities of *PvuII* structural neighbors. It is now clear that the catalytic domain in this group of enzymes, found in combination with various dimerization domains (Kovall and Matthews 1999), is related to a wide variety of other nucleic acid binding proteins including DNA repair endonucleases Vsr and MutH, 5'-3' λ -exonuclease (involved in recombination and repair), some Holliday junction resolvases and t-RNA exonuclease (Aravind et al. 2000; Declais et al. 2001). In addition, the type II restriction enzyme *NaeI* also exhibits recombinase activity (Jo and Topal 1995). Because PrfA has been implicated in DNA repair and recombination (Fernandez et al. 1998), it seems entirely reasonable that this protein is a member of the type II restriction enzyme superfamily.

Model building

Previous work (e.g., Rigden and Carneiro 1999; Rigden et al. 2000) has established a rigorous modeling methodology suitable for cases in which the target and template share low sequence identity. The threading-derived alignment is taken as a starting point to be analyzed, through extrapolation to multiple 3D models, and modified in the light of the results of protein structure validation tools.

The course of construction of the *B. subtilis* PrfA model is summarized in Figure 3. At stage 1, models were constructed based on a CLUSTAL W alignment, modified according to the 3D-PSSM results. Stages 2–6 represent models constructed during a series of alignment shifts. Examination of regions giving poor profile results was used to suggest local alignment modifications that would, for example, expose previously buried charged residues to solvent. Where the new set of models scored better than the existing ones, the alignment shift was accepted. In Figure 1 the alignment of PrfAs and *PvuII* is the best achieved. At

stages 7–9 “breeding” from the best model was carried out to improve stereochemical parameters. Residues in disallowed and generously allowed areas of the Ramachandran plot were subjected to manual intervention or remodeling so that they occupied more typical areas. Stage 0 in Figure 3 shows, for comparison, analysis of models produced using a CLUSTAL W derived alignment.

One notable finding is the poor quality of earlier models. The pG value (Sanchez and Sali 1998) of the best scoring model in stage 1 was just 0.46, below the threshold of 0.5 above which the model is considered reliable in terms of template choice and alignment quality. Even after two favorable alignment modifications the best scoring model scored just 0.52. The pG value corresponding to an average PROSA II score only comfortably exceeded the confidence threshold at stage 5, after four iterations of alignment improvement. Nevertheless, after the final alignment improvement at stage 6, the average pG value of the set of models was 0.93, close to the possible maximum of 1.0 and indicative of highly reliable models.

Although PROSA II was used as the main guide during the alignment modification process, VERIFY_3D scores for the model sets showed a clear upward trend (Fig. 3). Because the two validation methods are entirely independent, confidence that the changes in alignment represented true improvements in model quality was raised. The final model of *B. subtilis* PrfA contained no residues in disallowed regions of the Ramachandran plot and just three per subunit in the generously allowed areas. It has a good overall steric G-factor of -0.34 .

The final *B. subtilis* PrfA model was used as a basis to construct models of *B. stearothermophilus* PrfA. The backbones of the two sets of models were essentially identical, with just a single, readily accommodated two-residue insertion relative to *B. subtilis* PrfA needing to be modeled in the *B. stearothermophilus* protein. The *B. stearothermophilus* PrfA models were of similar quality to those of the *B. subtilis* protein with the best scoring -6.54 by PROSA II analysis, corresponding to a pG value of 0.97. These results confirm the expected compatibility of the *B. stearothermophilus* sequence with the structure of the *B. subtilis* PrfA model.

Analysis of the final *B. subtilis* PrfA model

During the process of model construction the assumption was made that PrfA was a dimer. This would be in accord with the large majority of known structures in the type II restriction enzyme superfamily, including *PvuII*. Size-exclusion chromatography data have suggested that at very low protein concentrations (<0.01 mg/mL) PrfA is monomeric in solution; however, at higher protein concentration (~ 1 mg/mL) the same chromatography experiments show a dimeric protein (Kelly et al. 2000). To resolve this question,

sedimentation velocity band and boundary analysis were performed, and they clearly showed a dimeric solution structure for *B. subtilis* PrfA (Table 1). Using the final model, additional support for the notion of a PrfA dimer comes from comparing analyses of dimeric and monomeric structures. Although the dimer structure is ranked first compared to the same sequence threaded into many decoy structures, for both the pair and solvent potentials by PROSA II analysis, the monomer ranks only 16th by pair potential and fourth by solvent potential. These data show that the interactions at the dimer interface of the model are favorable, and that exposure of interface residues to solvent would be unfavorable.

Although the DNA component of 1eyu was not included in the model building process, the final model, when superimposed on 1eyu, has no serious clashes with the DNA. The highly basic nature of the *B. subtilis* and *B. stearothermophilus* PrfA proteins (pI ~10) would be in general accord with a DNA binding capability. Possible DNA binding residues were therefore determined. Because the lack of DNA would possibly influence their conformation, manual side-chain readjustments of potential binding residues were permitted, but only to near-rotamer conformations. Eight positively charged residues potentially capable of electrostatically interacting with phosphate groups of a DNA backbone were found in this way (Fig. 4), along with four other resi-

dues capable of forming uncharged hydrogen bonds with the DNA backbone. Among the eight potential salt bridge forming residues, half are completely conserved as Arg or Lys when the phage and more distantly related *B. halodurans* sequences are not considered. Against a background of high sequence variability (Figs. 1–2) this is a significant finding that supports a functional role for these residues. Furthermore, just two of these 12 potential DNA binding residues differed in the *B. subtilis* and *B. stearothermophilus* models, suggesting that functional data obtained for one of the proteins can readily be extrapolated to the other.

The *PvuII* restriction enzyme has a well-defined double-stranded DNA sequence specificity, 5'-CAGCTG-3'. In contrast, little is known about the precise role of PrfA in vivo, and therefore, what its substrate might be. Two residues in the PrfA model seem suitably positioned to form specificity-determining hydrogen bonds to the DNA bases, Asn 115 and His 117, but what, if any, DNA sequence specificity PrfA has must remain a topic for future research.

The phenotypic implication of PrfA in both recombination and cell wall synthesis is unusual, and raises the possibility of dual substrate specificity, or even twin active sites. The latter possibility seems less likely because the positions of the MEME motifs discovered for this diverse sequence family cover regions of structural or functional importance in the model of PrfA as a dimeric DNA endonuclease (see above). No additional conserved regions, possibly indicative of a second catalytic site, are seen. One possibility might be that PrfA also displays activity against the polysaccharide chains comprising bacterial cell walls, that is, that the nuclease catalytic machinery also functions as a polysaccharidase. Although tentative modeling of polysaccharide chains into the PrfA model suggests no steric barriers to binding (data not shown), an indirect role for PrfA in cell wall synthesis, as suggested previously, may be more probable (Pedersen and Setlow 2000).

Fig. 4. Molscript (Kraulis 1991) diagram of the final model superimposed on the DNA from PDB entry 1eyu. One subunit is shown as a ribbon diagram, the other as a simple C α trace to better display key residues. The DNA is shown as a green double helix, with the exception of a segment marked with dotted lines. Side chains of presumed catalytic residues Asp 88, Asp 99, and Glu 101 (with yellow bonds) and possible specificity determinants Asn 115 and His 117 are marked on the lower subunit. Basic side chains well positioned to electrostatically interact with the phosphate groups of the DNA are marked on the upper subunit.

Nuclease activity of PrfA

The prediction from the model building that PrfA of *Bacillus* species could be a DNA binding protein has been partly tested previously, as *B. subtilis* PrfA has been shown to lack helicase activity (L. Pedersen and P. Setlow, unpubl.). The distant homology suggested to exist between PrfA and type II restriction enzymes is no guarantee of related catalytic activities (Devos and Valencia 2000). Nevertheless, the conservation of key catalytic site residues between the two families was suggestive of possible PrfA nuclease activity. This was first tested by analyzing the ability of purified *B. stearothermophilus* PrfA to cleave plasmid pUC19. These assays were carried out with the protein from *B. stearothermophilus*, because this organism has a temperature optimum for growth (60–65°C) very different from that of the *Escherichia coli* in which the PrfA was heterologously ex-

pressed. Assay at high temperature should therefore help eliminate nuclease activity originating from *E. coli*-derived impurities in the purified PrfA. The modeling (see above) and 60% sequence identity shared by the *B. subtilis* and *B. stearothermophilus* proteins (Fig. 1) ensured that functional inferences drawn for one of the proteins can be confidently extrapolated to the other (Devos and Valencia 2000).

B. stearothermophilus PrfA exhibited nuclease activity on DNA at 65°C, generating single-strand breaks in supercoiled plasmid thereby converting it into nicked circles (Fig. 5A, lanes 1–3; note also the small amount of linear DNA generated). Because the nicked circular plasmid DNA generated by PrfA could largely be converted to relaxed covalently closed circular DNA by phage T4 DNA ligase (Fig. 5A, lanes 4–5), the endonuclease activity of PrfA generates nicks with 5'-phosphate and 3'-hydroxyl termini. Complete nicking of supercoiled plasmid DNA in ~30 min required an excess of PrfA molecules over molecules of plasmid DNA (Fig. 5B). It was thus of significant concern that the low nuclease activity of the PrfA preparation was actually due to an *E. coli* protein that had copurified with PrfA. However, two separate experiments strongly indicate that this is unlikely. First, no nicked plasmid DNA was generated by the purified D86A PrfA variant in which the conserved aspartate residue (Fig. 1) likely to play a key role in any PrfA catalysis was mutated to alanine (Fig. 5C, compare lanes 2–3 with lanes 4–5). Second, the temperature optimum of the endonuclease activity of PrfA was 65°C (Fig. 6), which is common for proteins from the thermophile *B. stearothermophilus*. A significant amount of linear DNA was generated by wild-type PrfA at 24 to 45°C (Fig. 6, lanes 1–3) where any *B. stearothermophilus* protein would be expected to have low activity. However, at higher temperatures the amount of linear product was significantly lower, and nicked circular plasmid was by far the predominate product (Fig. 6, lanes 4–6). These data suggest that the endonuclease activity generating the nicked circular plasmid at elevated temperatures is due to the *B. stearothermophilus* PrfA. However, the generation of linear molecules, in particular at lower temperatures, could be due to *E. coli* contaminants in the recombinant PrfA preparation; note also that there was some generation of linear molecules by the D86A PrfA variant even at an elevated temperature (Fig. 5C, lanes 4–5). Generation of nicked circular plasmid by *B. stearothermophilus* PrfA was inhibited by EDTA, and was not stimulated by ATP, and generation of nicked DNA molecules was also seen when supercoiled plasmid pUB110 was used as substrate (data not shown). Other assays showed that the enzyme had no detectable exonuclease activity, and that purified *B. subtilis* PrfA was also able to nick supercoiled plasmid, but with a temperature optimum of ~37°C (data not shown).

Electrophoresis of PrfA-cleaved plasmid pUC19 on a denaturing agarose gel followed by Southern blot analysis

indicated that there was primarily only a single nick in the strand cleaved in the supercoiled DNA, as the great majority of the nicked strand was still full length (Fig. 7A, lanes 1,2). However, there were small amounts of DNA strands with multiple nicks, as hybridizing DNA fragments smaller than full length were detected; these smaller fragments were not detected in DNA treated with the D86A PrfA variant (Fig. 7A, lanes 2–3). Cleavage of PrfA-nicked pUC19 DNA with *Bam*H1 and Southern blot analysis of the DNA run on denaturing agarose gel electrophoresis showed that the sites of PrfA nicking were not random, as a number of discrete bands were detected in this analysis (Fig. 7A, lane 4). These discrete bands were not seen with plasmid incubated with the D86A PrfA variant followed by digestion with *Bam*H1 (data not shown). We have not yet identified the preferred sites of PrfA cleavage in pUC19 DNA, but ultimately this knowledge may give us some new insight into the specific function of PrfA in vivo.

Although PrfA clearly had endonuclease activity on supercoiled DNA, it was of obvious interest to determine if other types of DNAs were substrates for this enzyme. Strikingly, PrfA had very little endonuclease activity on relaxed, covalently closed plasmid DNA, as the specific activity on this substrate was at least 10-fold lower than that on supercoiled plasmid (Fig. 7B, lanes 1–6). Denaturing agarose gel electrophoresis and Southern blot analysis showed that *Bam*H1 linearized double-stranded DNA was cleaved by PrfA, and the sizes of the hybridizing bands detected in this analysis were identical to those obtained after *Bam*H1 digestion of PrfA nicked supercoiled plasmid DNA, and again, the D86A PrfA variant gave little if any cleavage of this DNA (Fig. 7A, lanes 4,5,6). Although PrfA exhibited similar specificity in its cleavage of supercoiled and linear DNAs, the enzyme was ~10-fold more active on the supercoiled template. The specific activity of enzyme on single-stranded DNA was also ≥20-fold lower than that on supercoiled DNA, as assessed using a closed circular single-stranded DNA as a substrate (Fig. 7B, lanes 3,4; Fig. 7C).

Although the DNA nicking activity of PrfA was very low, as noted above, the precise DNA substrate for PrfA in vivo is not known, and will almost certainly be different from the protein free supercoiled plasmid DNA that was the best substrate in our in vitro assays. Given the obvious potential involvement of proteins with the ability to cleave one strand of a DNA duplex in recombinational processes, it appears likely that the endonuclease activity observed for PrfA in vitro is a reflection of the enzymatic function of PrfA in vivo. The challenge now will be to determine the DNA substrate for PrfA in vivo, and the reaction catalyzed by the protein. Some clues that may guide this determination may be obtained by consideration of enzymes with endonuclease activity known to be involved in recombination processes.

One type of enzyme that plays a key role in the process of recombination is an endonuclease that can cleave Holler-

Fig. 5. Endonuclease activity of *B. stearotherophilus* PrfA: (A) products and specificity; (B) amount of protein required; and (C) lack of activity of a PrfA variant. (A) Incubation of supercoiled or relaxed plasmid pUC19 DNA with PrfA (0.5 μ g for 30 min at 65°C) was as described in Materials and Methods. Nicked circular pUC19 DNA (0.5 μ g) generated by PrfA cleavage at 65°C was isolated by agarose gel electrophoresis and treated with DNA ligase as described in Materials and Methods. Aliquots of various DNAs (~0.3–1 μ g) were run on agarose gel electrophoresis in the presence of 2 μ g/mL chloroquine as described in Materials and Methods. Identical conclusions were drawn based on a gel run with 4 μ g/mL chloroquine (data not shown). The samples in the various lanes were: lane 1, untreated DNA; lane 2, PrfA treated DNA; lane 3, topoisomerase treated DNA; lane 4, isolated nicked circular DNA generated by PrfA cleavage at 65°C; and lane 5, isolated nicked circular DNA generated by PrfA cleavage and then treated with phage T4 DNA ligase and ATP. The arrows labeled a, b, c, and d indicate the migration positions of nicked circular plasmid monomers, linear plasmid monomers, supercoiled monomeric plasmid topoisomers, and relaxed covalently closed plasmid monomers, respectively. The upper region of this gel containing the analogous dimeric plasmid species has been removed for clarity. (B) Supercoiled plasmid pUC19 was incubated with varying amounts of PrfA for 1 h at 65°C and the DNA run on agarose gel electrophoresis as described in Materials and Methods. The amounts of PrfA in the samples run in the lanes were: lane 1, none; lane 2, 0.5 μ g; lane 3, 0.25 μ g; lane 4, 0.125 μ g; lane 5, 0.063 μ g; and lane 6, 0.031 μ g. The arrows labeled a, b, c, and d indicate the migration positions of supercoiled dimeric plasmid, nicked circular monomeric plasmid, linear monomeric plasmid, and supercoiled monomeric plasmid, respectively. The region of this gel above the supercoiled dimeric plasmid has been removed for clarity. (C) Supercoiled plasmid pUC19 DNA was incubated with varying amounts of wild-type PrfA or the D86A variant for 1 h at 65°C, and aliquots run on agarose gel electrophoresis as described in Materials and Methods. The amounts and types of PrfA in the samples run in the various lanes were: lane 1, no PrfA; lane 2, 0.5 μ g wild-type PrfA; lane 3, 2.5 μ g wild-type PrfA; lane 4, 0.5 μ g D86A PrfA; lane 5, 2.5 μ g D86A PrfA. The arrows labeled a, b, c, and d indicate the migration positions of supercoiled dimeric plasmid, nicked circular monomeric plasmid, linear monomeric plasmid, and supercoiled monomeric plasmid, respectively. The region of this gel above the supercoiled dimeric plasmid species has been removed for clarity.

Fig. 6. Temperature optimum of the endonuclease activity of *B. stearotherophilus* PrfA on DNA. Supercoiled plasmid pUC19 was incubated with or without purified *B. stearotherophilus* PrfA (0.5 μ g) for 1 h at various temperatures and the reactions analyzed by agarose gel electrophoresis as described in Materials and Methods. The samples in the various lanes are labeled at the top of the gel and were incubated at: lane 1, 24°C; lane 2, 37°C; lane 3, 45°C; lane 5, 55°C; lane 6, 65°C; lane 7, 75°C; and lane 8, 75°C, with no PrfA. Lane 4 contains molecular weight markers of 1.9, 2.3, 4.5, 6.3, 9.7, and 19.7 kb, reading from the bottom to the top of the gel. The labeled arrows denote the migration positions of: (a) relaxed dimeric plasmid; (b) linear dimeric plasmid; (c) supercoiled dimeric plasmid; (d) relaxed monomeric plasmid; (e) linear dimeric plasmid; and (f) supercoiled monomeric plasmid. Plasmid before incubation gave a pattern on agarose gel electrophoresis that was identical to that seen in lane 8.

day junctions, which are key intermediates in recombination (Declais et al. 2001; Lilley and White 2001; Sharples 2001). These enzymes are invariably dimers and some exhibit significant structural homology to type II restriction enzymes. In vitro, these endonucleases often exhibit significant sequence specificity (Lilley and White 2001; Sharples 2001), and cleave cruciform structures formed and stabilized in supercoiled DNA, but generally have much lower activity on relaxed circular or linear DNA. Normally, these enzymes cleave both DNA strands comprising a cruciform structure, but it is not inconceivable that under some conditions or with some templates the protein dissociates from the now relaxed template so rapidly after the first cleavage event that the second cleavage does not take place. The plasmid template routinely used as the substrate has not been engineered to contain a cruciform structure, but plasmid supercoiling might well facilitate the transient formation of such structures. However, plasmid pUC19 is certainly not the true substrate for PrfA in vivo, as is evidenced by its extremely slow cleavage by PrfA. As suggested above, use of such a poor substrate could also contribute significantly to the extremely slow cleavage of the second strand in any very transient cruciform structure.

Homologous recombination between linear eukaryotic chromosomes generates linear products that have a structure similar to their parents, and the recombination process does not represent a problem for partitioning. However, in the

Fig. 7. Action of PrfA on various DNA substrates. (A) Action and specificity of PrfA on supercoiled and linear DNA. PrfA was incubated with various forms of plasmid pUC19 DNA for 30 min at 65°C, and the DNA denatured, run on agarose gel electrophoresis and subjected to Southern blot analysis using a plasmid pUC19 probe as described in Materials and Methods. The samples run in the various lanes were: lane 1, supercoiled DNA with no PrfA; lane 2, supercoiled DNA treated with 0.2 μ g PrfA; lane 3, supercoiled DNA treated with 0.25 μ g D86A PrfA; lane 4, supercoiled DNA treated with 0.25 μ g PrfA, and cleaved with *Bam*H1 after PrfA treatment; lane 5, *Bam*H1 linearized DNA treated with 2.5 μ g PrfA; and lane 6, *Bam*H1 linearized DNA treated with 2.5 μ g D86A PrfA. The letters adjacent to lanes in the figure denote the migration positions of: (a) supercoiled dimeric plasmid; (b) linear single-stranded monomeric plasmid; (c) single-stranded circular monomeric plasmid; and (d) supercoiled monomeric plasmid that migrates in this system at a size of $\sim$ 1.6 kb. Note that the supercoiled monomeric plasmid is very poorly detected in this system, as expected. The region of the gel above the migration position of supercoiled dimeric plasmid (lanes 1–3) or linear monomeric plasmid (lanes 4–6) has been removed for clarity. (B) Action of PrfA on supercoiled and relaxed plasmid DNA. Supercoiled or topoisomerase-relaxed plasmid pUC19 DNA was incubated with PrfA for 30 min at 65°C, and DNA run on agarose gel electrophoresis in the presence of chloroquine (2 μ g/mL) as described in Materials and Methods. The samples run in the various lanes were: lane 1, supercoiled plasmid, no PrfA; lane 2, relaxed plasmid, no PrfA; lane 3, supercoiled plasmid, 0.15 μ g PrfA; lane 4, supercoiled plasmid, 0.3 μ g PrfA; lane 5, relaxed plasmid, 0.3 μ g PrfA; and lane 6, relaxed plasmid, 0.6 μ g PrfA. The letters to the left of the figure denote the migration positions of: (a) nicked monomeric plasmid; (b) linear monomeric plasmid; (c) supercoiled monomeric plasmid topoisomers; and (d) relaxed monomeric plasmid. The region of the gel above the migration position of nicked monomeric plasmid has been removed for clarity. (C) Action of PrfA on circular single-stranded DNA. Single-stranded circular M13 DNA (1 μ g) was incubated for 2 and 1/2 h at 65°C with PrfA or the D86A PrfA variant, and the DNA run on agarose gel electrophoresis as described in Materials and Methods. The samples in the various lanes contained: lane 1, no additions; lane 2, 0.25 μ g PrfA; and lane 3, 0.25 μ g D86A PrfA. The letters a and b to the left of the figure denote the migration positions of circular and linear single-stranded DNA, respectively, and the letters c and d to the right of the figure denote the migration positions of 6.7 and 4.5-kb double-stranded linear DNA markers, respectively. The DNA without incubation looked exactly like the DNA in lane 1 (data not shown).

case of bacterial chromosomes, their circularity represents a potential impediment to the maintenance of their integrity (Leslie and Sherratt 1995). Any odd number of homologous recombinational events generates a fusion between the two daughter chromosomes, called chromosome dimers, and chromosome dimers cannot be properly segregated into the two daughter cells at cell division. However, a site-specific recombination system has been described, which operates at the replication terminus. In *E. coli*, this system consists of *dif*, a 28-bp DNA segment located in the terminus, and XerC and XerD, two site-specific recombinases that act at *dif* to ensure resolution of dimeric chromosomes into monomers (Blakely et al. 1997). Only monomeric chromosomes can be stably inherited. Such a site-specific recombination system has also recently been described in *B. subtilis*: *dif* is the DNA target site where CodV and RipX, two site-specific recombinases, act (Sciocchetti et al. 2001).

Defects in the *dif*/XerC/XerD or *dif*/CodV/RipX site-specific recombination systems generate what is known as the *dif* phenotype, part of which is manifested as an intersecting of the centrally located nucleoid by the forming partitioning septum (Hendricks et al. 2000; Sciocchetti et al. 2001) in about 10–15% of the cells. The latter are cells where the two daughter chromosomes remained dimeric as a consequence

of homologous recombination and were not separated by the defective site-specific recombination system.

It is interesting to note that in a *B. subtilis* *prfA* mutant, one of the abnormalities that have been described is the presence of nucleoids that appear to be intersected by the septum (Pedersen and Setlow 2000), suggesting a chromosome segregation defect. It has also been noted that in *B. subtilis* a *recA* mutation does not suppress the *ripX* phenotype (Sciocchetti et al. 2001). This contrasts with the situation in *E. coli*, where *recA* mutations suppress a *xerC* or *xerD* phenotype. This suggests that there is another activity in *B. subtilis* that performs abundant homologous recombination in the absence of *recA*. Interestingly, *B. subtilis* and other Gram-positive bacteria lack recognizable homologs of Holliday junction resolving endonucleases (Sharples et al. 1999; Sharples 2001). Given these latter two observations, it is tempting to speculate that PrfA is this missing protein involved in homologous recombination in Gram-positive bacteria, but this remains a subject for further work.

Another protein, FtsK, has been shown to be involved in resolution at the *dif* sequence and in chromosome partitioning (Recchia et al. 1999). It was suggested that this protein localizes at the division septum and ensures that chromosome dimers are converted to monomers. FtsK is targeted to

the FtsZ ring at the division septum via its N-terminal region. An attractive hypothesis regarding the function of PrfA is that this protein has a role similar to the one performed by FtsK, as previously suggested (Pedersen and Setlow 2000). It is therefore notable that PBP-1 (penicillin-binding protein 1) localizes to sites of cell division (Pedersen et al. 1999), because PBP-1 and PrfA have already been linked phenotypically and genetically. The exacerbation of the *prfA* phenotype by the additional loss of PBP-1 (a *ponA* mutation) but not by the loss of other penicillin-binding proteins, suggests that PrfA and PBP-1 might be involved in some common cellular function; this is, of course, consistent with the cotranscription of *prfA* and *ponA* in an operon (Popham and Setlow 1995).

The endonuclease activity that has been described in this work for PrfA exhibits a number of similarities to the activities of endonuclease involved in recombination plus the recombination and the chromosomal segregation defects in *prfA* mutant cells strongly suggest that PrfA is involved directly in recombination and/or possibly chromosome segregation. The presence of *prfA*-related genes in so many organisms and their localization in an operon with penicillin-binding protein encoding genes in many organisms is further suggestive of PrfA participation in a process that is present in all these bacteria.

Conclusions

The work presented here provides a further example of the additional insights that may be revealed through the use of fold recognition programs in cases where simple sequence comparisons are inadequate. Although the number of genomic sequences continues to grow enormously and structural genomics programs are still in their infancy (Thornton 2001), the need to extract the maximum amount of information from sequences through fold recognition and modeling remains high. The challenge now lies in automation of these steps. The present work and previous studies identify several large barriers to such automation. Among them, the comparison of biochemical data for the target and characteristics of the putative templates, so important in supporting initial fold recognition methods, remains difficult, although some positive steps have recently been made (MacCallum et al. 2000; Ono et al. 2001). A second key problem is that the initial alignments produced by threading programs for correct templates may, as here, not be of sufficient quality for the production of objectively reliable models. Automated mechanisms for improvement of threading-derived models (Kolinski et al. 1999) can therefore only grow in importance. Nevertheless, whether automated or manual, studies such as this and others (e.g., Rigden et al. 2001) in which modeling correctly predicts function should encourage further methodological development.

The major conclusions derived from this work are thus that PrfA is structurally related to the type II restriction enzyme superfamily and that PrfA has an endonuclease activity that is reminiscent of that of enzymes involved in recombinational processes. Although these conclusions in and of themselves do not establish a specific role for PrfA in vivo, they are certainly consistent with a role for this protein in DNA recombination or possibly chromosome segregation, and suggest further experimental work that may delineate the specific role for PrfA in these processes.

Materials and methods

Sequence analysis

A search of the nr database for sequences homologous to *B. subtilis* PrfA was made using HMM methods (Karplus et al. 1998). Searches were also carried out with Position-Specific Iterated BLAST (PSI-BLAST) (Altschul et al. 1997). Percentage pairwise identities among the resultant sequence set were calculated using MODELLER-4 (Sali and Blundell 1993) and a neighbor-joining tree representation of their sequence relationships produced using programs of the PHYLIP package (Felsenstein 1989). Conserved sequence motifs within the group of PrfA homologs were obtained using the Multiple EM for Motif Elucidation (MEME) software (Grundy et al. 1996). The program Jalview (available at <http://circinus.ebi.ac.uk:6543/jalview/>) was used to manipulate alignments. Secondary structure predictions were made by the PHD (Rost 1996), Hidden Markov Model (HMM) (Karplus et al. 1998), and PSIPRED (Jones 1999) methods.

Fold recognition experiments

Methods for fold recognition are not equally successful for a given sequence; success may be obtained for a given sequence with one method but not with another, while a different sequence may produce the opposite result. Fold recognition experiments were therefore carried out with several of the more successful Web-based services: 3D-PSSM, Genthreader, FFAS, and the Bioinbgu suite of related methods (Fischer et al. 1999).

Model building and validation

Several structures of *PvuII* have been deposited in the PDB, including both wild-type and mutant proteins, with and without DNA and with and without bound metal ions. The highest resolution wild-type *PvuII* structure, 1eyu (Horton and Cheng 2000), was chosen as the template for model construction.

The limited sequence identity between *B. subtilis* PrfA and *PvuII* (around 15%) necessitated the application of a rigorous modeling strategy in which construction and evaluation of multiple models was used to validate the target-templates alignment (e.g., Rigden and Carneiro 1999; Rigden et al. 2000). In this way, 15 models were built for each target-template alignment tested with PrfA. Protein models were constructed using the MODELLER-4 package (Sali and Blundell 1993). A 4-Å coordinate randomization step was applied prior to refinement of the models to sample coordinate space. The set of models was then analyzed for solvent exposure and packing with PROSA II (Sippl 1993) and VERIFY_3D (Lüthy et al. 1992). Both of these programs produce

an overall score for a given structure, with correct structures yielding values within a characteristic size-dependent range and profiles highlighting regions of dubious solvent exposure and packing characteristics (Lüthy et al. 1992; Sippl 1993). Possible problematic zones are characterized by negative VERIFY_3D profiles or strongly positive PROSA II profiles. In conjunction with the model length, PROSA II scores were used to calculate pG values (Sanchez and Sali 1998). These fall in the range 0–1, with values above 0.5 taken to indicate correct choice of template and largely correct alignment. Stereochemical properties were analyzed with PROCHECK (Laskowski et al. 1993). Structural superpositions were made with LSQMAN (Kleywegt 1996), and structures were visualized with O (Jones et al. 1991).

Production of B. subtilis and B. stearothermophilus PrfA

The *B. subtilis* and *B. stearothermophilus prfA* genes have been cloned into an *E. coli* expression strain, and the resulting overexpressing proteins were purified to homogeneity as previously described (Kelly et al. 2000; Pedersen and Setlow 2000). The purity of the final purified PrfA proteins was greater than 98%, as judged by SDS-PAGE analysis. The purified proteins migrated on SDS-PAGE as single bands of ~25 kD (data not shown), corresponding to the calculated molecular weights (based on the sequence) of 23,959 and 23,034 D for *B. subtilis* and *B. stearothermophilus* PrfAs, respectively.

Site-directed mutagenesis of B. stearothermophilus prfA

Site-directed mutagenesis of the *B. stearothermophilus prfA* gene was carried out by PCR. The upstream primer used was 5'-CAT **GCCATGG**CACTCAAATACCCGAG, in which the underlined ATG is the PrfA translation start codon that is also a part of an *NcoI* site (bold-faced residues). The downstream primer was chosen to change PrfA residue 86 from a conserved Asp residue (Fig. 1) to Ala. The sequence of this primer was 5'GTACTTGCC **GCGGTACACACCGTTGTAGGCGGTCGTCG**; the nucleotide changed is underlined, and a *SacII* site in this primer also present in the *prfA* gene is in bold face. PCR-using plasmid pPS3248 (Kelly et al. 2000) carrying the *B. stearothermophilus prfA* gene as a template with these two primers gave the expected 283-bp fragment that was cloned in plasmid pCR2.1 (Invitrogen); one plasmid with the expected insert was identified by DNA sequence analysis, and was named pPS3464. A ~270-bp fragment was excised from this plasmid by digestion with *NcoI* and *SacII* and cloned between the *NcoI* and *SacII* sites in plasmid pPS3248 (after removal of the wild-type 270-bp *NcoI*-*SacII* fragment) in *E. coli* TG1 giving strain PS3465. The plasmid in this strain has the mutated *prfA* gene inserted in plasmid pET9d (Kelly et al. 2000). Prior to expression of the mutant PrfA, the plasmid from strain PS3465 was transformed into *E. coli* BL21 (DE3 pLysS) giving strain PS3466, which was the strain used for overexpression of the D86A PrfA variant. The mutant PrfA was overexpressed and purified as described above, and behaved identically to wild-type PrfA throughout purification.

Hydrodynamic characterization

Analytical ultracentrifugation band and boundary sedimentation velocity measurements were performed using a Beckman XLA

analytical ultracentrifuge on *B. subtilis* PrfA protein concentrated to 1 mg/mL in 10 mM Na phosphate buffer (pH 6.0). Radial scanning was performed at 280 nm. Band sedimentation was performed as described (Jedrzejewski et al. 2000). The correction values for 90% D₂O were 1.005 for the relative viscosity and 1.0103 for the buoyancy term (Kirschenbaum 1951). These experiments were performed at three NaCl concentrations, 50, 100, and 150 mM. Boundary experiments were analyzed using time-derived software (Stafford 1992), and the partial specific volume, v , was calculated based on the PrfA sequence. A standard formula was used to correct the s values to $s_{20,w}$ (van Holde 1985). The diffusion coefficient and the molecular weight were calculated using Sedband (P. Schuck, unpubl.).

Nuclease assays

The nuclease activity of purified PrfA was routinely assayed with plasmid pUC19 containing ~75% supercoiled forms. Reactions were routinely carried out in 10 μ L of 25 mM HEPES (pH 7.5) and 1 mM MgCl₂ with 1.2 μ g of plasmid pUC19 DNA. After incubation at various temperatures, reactions were cooled on ice, immediately run on agarose gel electrophoresis, and the gel stained with ethidium bromide. In some cases reactions were run on agarose gels containing chloroquine (2 μ g/mL and 4 μ g/mL) as described (Nicholson et al. 1990) to resolve plasmid topoisomers, as well as nicked circular and linear plasmid DNA. Relaxed closed circular plasmid DNA was prepared by treatment of supercoiled plasmid with calf topoisomerase I as described (Nicholson et al. 1990), and linear plasmid DNA was prepared by digestion with *Bam*H1. Plasmid DNA before or after PrfA treatment was also denatured with glyoxal, run on agarose gel electrophoresis, DNA transferred to nitrocellulose-based paper and the plasmid DNA on the paper detected by hybridization with a pUC19 probe as described (Fairhead et al. 1993). Covalently closed circular single-stranded M13mp18 viral DNA was purchased from New England Biolabs, and conversion of this DNA to a linear molecule was also monitored by agarose gel electrophoresis.

Other methods

Electrophoresis was carried out under reducing conditions in 12% polyacrylamide gels using the buffer system described by Laemmli (1970) and a Mini Protein II gel system (Bio-Rad). Protein concentration was determined by UV absorption at 280 nm using calculated molar extinction coefficients (Pace et al. 1995).

Acknowledgments

The authors thank Tobie Bogart for her assistance with the sedimentation data collection. Work in the authors' laboratories was supported by grants AI44079 (M.J.J.) and GM19698 (P.S.) from the National Institutes of Health, and N66001-01-C-8013 from the Defense Advanced Research Projects Agency (M.J.J. and P.S.).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.

- Aravind, L., Makarova, K.S., and Koonin, E.V. 2000. Holliday junction resolvases and related nucleases: Identification of new families, phyletic distribution and evolutionary trajectories. *Nucleic Acids Res.* **28**: 3417–3432.
- Ban, C. and Yang, W. 1998. Structural basis for MutH activation in *E. coli* mismatch repair and relationship of MutH to restriction enzymes. *EMBO J.* **17**: 1526–1534.
- Barton, G.J. 1993. ALSCRIPT, a tool to format multiple sequence alignments. *Protein Eng.* **6**: 37–40.
- Blakely, G.W., Davidson, A.O., and Sherratt, D.J. 1997. Binding and cleavage of nicked substrates by site-specific recombinases XerC and XerD. *J. Mol. Biol.* **265**: 30–39.
- Cheng, X., Balendiran, K., Schildkraut, I., and Anderson, J. E. 1994. Structure of PvuII endonuclease with cognate DNA. *EMBO J.* **13**: 3927–3935.
- Declais, A.-C., Hadden, J., Phillips, S.E.V., and Lilley, D.M.J. 2001. The active site of the junction-resolving enzyme T7 endonuclease I. *J. Mol. Biol.* **307**: 1145–1158.
- Devos, D. and Valencia, A. 2000. Practical limits of function prediction. *Proteins* **41**: 98–107.
- Fairhead, H., Setlow, B., and Setlow, P. 1993. Prevention of DNA damage in spores and in vitro by small, acid-soluble proteins from *Bacillus* species. *J. Bacteriol.* **175**: 1367–1374.
- Felsenstein, J. 1989. PHYLIP—Phylogeny inference package (version 3.2). *Cladistics* **5**: 164–166.
- Fernandez, S., Sorokin, A., and Alonso, J.C. 1998. Genetic recombination in *Bacillus subtilis* 168: Effects of recU and recS mutations on DNA repair and homologous recombination. *J. Bacteriol.* **180**: 3405–3409.
- Fischer, D. 2000. Hybrid fold recognition: Combining sequence derived properties with evolutionary information. In *Pacific Symp. Biocomputing, Hawaii* (eds. R.B. Altman, A.K. Dunker, A. Hunter, K. Lauderdale, and T.E. Klein), pp. 119–130. World Scientific, Singapore.
- Fischer, D., Barret, C., Bryson, K., Eloffson, A., Godzik, A., Jones, D., Karplus, K.J., Kelley, K.A., MacCallum, R.M., Pawowski, K., Rost, B., Rychlewski, L., and Sternberg, M. 1999. CAFASP-1: Critical assessment of fully automated structure prediction methods. *Proteins (Suppl.)* **3**: 209–217.
- Ghuysen, J.M. 1994. Molecular structures of penicillin-binding proteins and β -lactamases. *Trends Microbiol.* **2**: 372–380.
- Grundy, W.N., Bailey, T.L., and Elkan, C.P. 1996. ParaMEME: A parallel implementation and a web interface for a DNA and protein motif discovery tool. *Comput. Appl. Biosci.* **12**: 303–310.
- Hendricks, E.C., Szerlong, H., Hill, T., and Kuempel, P. 2000. Cell division, guillotining of dimer chromosomes and SOS induction in resolution mutants (dif, xerC and xerD) of *Escherichia coli*. *Mol. Microbiol.* **36**: 973–981.
- Horton, J.R. and Cheng, X. 2000. PvuII endonuclease contains two calcium ions in active sites. *J. Mol. Biol.* **300**: 1049–1056.
- Jedrzejewski, M.J., Hollingshead, S.K., Lebowitz, J., Chantlat, L., Briles, D.E., and Lamani, E. 2000. Production and characterization of the functional fragment of pneumococcal surface protein A. *Arch. Biochem. Biophys.* **373**: 116–125.
- Jo, K. and Topal, M.D. 1995. DNA topoisomerase and recombinase activities in Nae I restriction endonuclease. *Science* **267**: 1817–1820.
- Jones, D.T. 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**: 195–202.
- Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.* **A47**: 110–119.
- Karplus, K., Barrett, C., and Hughey, R. 1998. Hidden Markov models for detecting remote protein homologies. *Bioinformatics* **14**: 846–856.
- Kelley, L.A., MacCallum, R.M., and Sternberg, M.J.E. 2000. Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.* **299**: 501–522.
- Kelly, S.J., Stein, R.A., Bagyan, I., Setlow, P., and Jedrzejewski, M.J. 2000. Structural characterization of penicillin-binding protein-related factor A (PrfA) from *Bacillus* species. *J. Struct. Biol.* **131**: 90–95.
- Kirschenbaum, I. 1951. *Physical properties and analysis of heavy water*. McGraw-Hill, New York, NY.
- Kleywegt, G.J. 1996. Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallogr.* **D52**: 842–857.
- Kolinski, A., Rotkiewicz, P., Ilkowsky, B., and Skolnick, J. 1999. A method for the improvement of threading-based protein models. *Proteins* **37**: 592–610.
- Kovall, R.A. and Matthews, B.W. 1999. Type II restriction endonucleases: Structural, functional and evolutionary relationships. *Curr. Opin. Chem. Biol.* **3**: 578–583.
- Kraulis, J. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**: 946–950.
- Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680–685.
- Laskowski, R., MacArthur, M., Moss, D., and Thornton, J. 1993. PROCHECK: A program to check stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**: 283–290.
- Leslie, N.R. and Sherratt, D.J. 1995. Site-specific recombination in the replication terminus region of *Escherichia coli*: Functional replacement of dif. *EMBO J.* **14**: 1561–1570.
- Lilley, D.M.J. and White, M.F. 2001. The junction-resolving enzymes. *Nat. Rev. Mol. Biol.* **2**: 433–443.
- Lüthy, R., Bowie, J.U., and Eisenberg, D. 1992. Assessment of protein models with three-dimensional profiles. *Nature* **356**: 83–85.
- MacCallum, R.M., Kelley, L.A., and Sternberg, M.J. 2000. SAWTED: Structure assignment with text description-enhanced detection of remote homologues with automated SWISS-PROT annotation comparison. *Bioinformatics* **16**: 125–129.
- Nicholson, W.L., Setlow, B., and Setlow, P. 1990. Binding of DNA in vitro by a small, acid-soluble spore protein from *Bacillus subtilis* and the effect of this binding on DNA topology. *J. Bacteriol.* **172**: 6900–6906.
- Ono, T., Hishigaki, H., Tanigami, A., and Takagi, T. 2001. Automated extraction of information on protein-protein interactions from the biological literature. *Bioinformatics* **17**: 155–161.
- Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. 1995. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4**: 2411–2423.
- Pedersen, L.B. and Setlow, P. 2000. Penicillin-binding protein-related factor A is required for proper chromosome segregation in *Bacillus subtilis*. *J. Bacteriol.* **182**: 1650–1658.
- Pedersen, L.B., Angert, E.R., and Setlow, P. 1999. Septal localization of penicillin-binding protein 1 in *Bacillus subtilis*. *J. Bacteriol.* **181**: 3201–3211.
- Pinho, M.G., de Lencastre, H., and Tomasz, A. 1998. Transcriptional analysis of the *Staphylococcus aureus* penicillin binding protein 2 gene. *J. Bacteriol.* **180**: 6077–6081.
- Popham, D.L. and Setlow, P. 1995. Cloning, nucleotide sequence and mutagenesis of the *Bacillus subtilis* *ponA* operon, which encodes for penicillin-binding protein (PBP) 1 and a PBP-related factor. *J. Bacteriol.* **177**: 326–335.
- Recchia, G.D., Aroyo, M., Wolf, D., Blakely, G., and Sherratt, D.J. 1999. FtsK-dependent and -independent pathways of Xer site-specific recombination. *EMBO J.* **18**: 5724–5734.
- Rigden, D.J. and Carneiro, M. 1999. A structural model for the *rolA* protein and its interaction with DNA. *Proteins* **37**: 697–708.
- Rigden, D.J., Mello, L.V., and Bertioli, D.J. 2000. Structural modeling of a plant disease resistance gene product domain. *Proteins* **41**: 133–143.
- Rigden, D.J., Bagyan, I., Lamani, E., Setlow, P., and Jedrzejewski, M.J. 2001. A cofactor-dependent phosphoglycerate mutase homologue from *Bacillus* species is actually a broad specificity acid phosphatase. *Protein Sci.* **10**: 1835–1846.
- Rost, B. 1996. PHD: Predicting one-dimensional protein structure by profile based neural methods. *Methods Enzymol.* **266**: 525–539.
- Rychlewski, L., Jaroszewski, L., Li, W., and Godzik, A. 2000. Comparison of sequence profiles. Strategies for structural predictions using sequence information. *Protein Sci.* **9**: 232–241.
- Sali, A. and Blundell, T.L. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**: 779–815.
- Sanchez, R. and Sali, A. 1998. Large scale protein structure modeling of the *Saccharomyces cerevisiae* genome. *Proc. Natl. Acad. Sci.* **95**: 13597–13602.
- Sciochetti, S.A., Piggot, P.J., and Blakely, G.W. 2001. Identification and characterization of the dif site from *Bacillus subtilis*. *J. Bacteriol.* **183**: 1058–1068.
- Sharples, G.J. 2001. The X philes: Structure-specific endonucleases that resolve Holliday junctions. *Mol. Microbiol.* **39**: 823–834.
- Sharples, G.J., Ingleston, S.M., and Lloyd, R.G. 1999. Holliday junction processing in bacteria: Insights from the evolutionary conservation of RuvABC, RecG, and RusA. *J. Bacteriol.* **181**: 5543–5550.
- Sippl, M.J. 1993. Recognition of errors in three-dimensional structures of proteins. *Proteins* **17**: 355–362.
- Stafford, W.F. 1992. Boundary analysis in sedimentation transport experiments: A procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile. *Anal. Biochem.* **203**: 295–301.
- Thornton, J.M. 2001. Structural genomics takes off. *Trends Biochem. Sci.* **26**: 88–89.
- van Holde, K.E. 1985. *Physical biochemistry*. Prentice Hall, Englewood Cliffs, NJ.
- Wake, R.G. and Errington, J. 1995. Chromosome partitioning in bacteria. *Annu. Rev. Genet.* **29**: 41–67.