Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Crystallization and preliminary X-ray diffraction analysis of SeqA bound to a pair of hemimethylated GATC sites

Yu Seon Chung and Alba Guarné*

Department of Biochemistry and Biomedical Sciences, HSC-4N57A, McMaster University, Hamilton ON L8N 3Z5, Canada

Correspondence e-mail: guarnea@mcmaster.ca

Received 17 March 2008
Accepted 16 May 2008

Escherichia coli SeqA is a negative regulator of DNA replication. The SeqA protein forms a high-affinity complex with newly replicated DNA at the origin of replication and thus prevents premature re-initiation events. Beyond the origin, SeqA is found at the replication forks, where it organizes newly replicated DNA into higher ordered structures. These two functions depend on SeqA binding to multiple hemimethylated GATC sequences. In an effort to understand how SeqA forms a high-affinity complex with hemimethylated DNA, a dimeric variant of SeqA was overproduced, purified and crystallized bound to a DNA duplex containing two hemimethylated GATC sites. The preliminary X-ray analysis of crystals diffracting to 3 Å resolution is presented here.

1. Introduction

All the GATC sequences within the *Escherichia coli* chromosome normally exist in their fully methylated state. Upon initiation of DNA replication, GATC sequences are transiently hemimethylated until the newly incorporated adenine is methylated by Dam methylase. Regulation of several important biological processes such as mismatch repair or timing of replication re-initiation depends on this period of hemimethylation (Barras & Marinus, 1989; Campbell & Kleckner, 1990; Modrich & Lahue, 1996). The SeqA protein binds clusters of hemimethylated GATC sites and thus prevents premature re-initiation events in a process known as origin (*oriC*) sequestration (Lu *et al.*, 1994). Upon initiation of replication, SeqA prevents binding of the main initiator protein DnaA to the low-affinity DnaA-binding boxes that include GATC sequences and hence SeqA sequestration resets the conformation of the newly replicated origin (Nievera *et al.*, 2006).

Interestingly, clusters of SeqA bound to DNA are detected at the replication forks rather than the origin. The formation of these foci depends on Dam methylation and ongoing replication, but does not require the presence of *oriC*, suggesting that SeqA has additional functional roles beyond sequestration (Hiraga *et al.*, 2000; Niki & Hiraga, 1998; Onogi *et al.*, 1999). Indeed, loss of *seqA* causes asynchrony of replication, hyper-initiation and abnormal localization of nucleoids (Bach & Skarstad, 2004; Boye *et al.*, 1996). Purified SeqA forms long and polydisperse linear polymers that are able to form a multivalent complex with DNA (Guarné *et al.*, 2005). Formation of these SeqA filaments on newly replicated DNA has been postulated to restrain negative supercoils on the DNA (Guarné *et al.*, 2005; Odsbu *et al.*, 2005). Conversely, mutants of SeqA that have lost the ability to oligomerize introduce positive supercoils onto DNA, revealing a rich pattern of interactions between SeqA and newly replicated DNA (Odsbu *et al.*, 2005).

The structure of the C-terminal domain of SeqA bound to a single hemimethylated GATC revealed the nature of the specific interactions with DNA and their dependence on methylation (Guarné *et al.*, 2002). However, SeqA only interacts tightly with DNA when it is bound to multiple hemimethylated GATC sequences (Brendler &

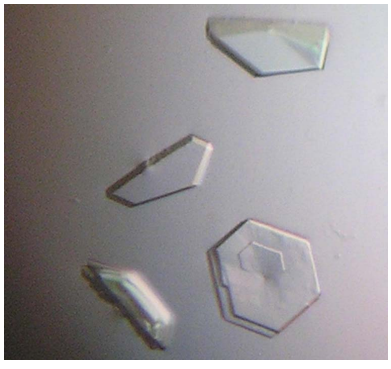

Table 1
Hemimethylated DNA duplexes crystallized with SeqA(Δ41–59)-A25R.

Methylated adenines are indicated in bold. CG snap-fasteners are shaded dark grey and GATC sequences light grey.

DNA sequence		Length/spacing	Crystallization conditions	Resolution (Å)
A	5′GAGTCG**A**TCGGTGCG**A**TCCTTAG3′ 3′CTCAGCTAGCCACGCTAGGAATC5′	23/9	20% PEG 1000, 0.1 *M* Tris pH 9	6
B	5′GAGTCG**A**TCGGTGCTG**A**TCCTTAG3′ 3′CTCAGCTAGCCACGACTAGGAATC5′	24/10	5–7% PEG 400, 0.15 *M* KCl, 0.01 *M* $MgCl_2$	12
C	5′GAGTCG**A**TCGGTGCTG**A**TCCTTA3′ 3′CTCAGCTAGCCACGACTAGGAAT5′	23/10	6.5% PEG 10 000, 0.25 *M* ammonium acetate, 0.1 *M* bis-Tris pH 6.5	7
D	5′GAGTCG**A**TCGGTGCG**A**TCCTTA3′ 3′TCAGCTAGCCACGCTAGGAATC5′	21/9	10% PEG 400, 0.1 *M* sodium citrate, 0.1 *M* Tris pH 7.5	5
E	5′GAGTCG**A**TCGGCGGG**A**TCCTTA3′ 3′TCAGCTAGCCGCCCTAGGAATC5′	21/9	10% PEG 400, 0.1 *M* sodium citrate, 0.1 *M* Tris pH 7.5	5
F	5′GAGTCG**A**TCGGCGGG**A**TCCTTA3′ 3′CAGCTAGCCGCCCTAGGAATCT5′	20/9	14–20% MPD, 0.3–0.4 *M* ammonium acetate, 0.1 *M* sodium citrate pH 5.6	3

Austin, 1999). Moreover, the balance between positive/negative supercoils introduced by SeqA onto newly replicated DNA also relies on its ability to recognize multiple hemimethylated GATC sites (Odsbu *et al.*, 2005). In an effort to understand the topological constraints imposed by SeqA on newly replicated DNA, we have undertaken the structure determination of a dimeric SeqA mutant bound to a pair of hemimethylated GATC sequences. Here, we report the optimization of hemimethylated DNA duplexes, their crystallization in complex with the SeqA(Δ41–59)-A25R mutant and the preliminary characterization of SeqA(Δ41–59)-A25R–DNA cocrystals.

2. Experimental and results

2.1. Cloning of SeqA(Δ41–59)-A25R

The *seqA* coding sequence was subcloned in pET-11a as previously described (Brendler & Austin, 1999). Wild-type SeqA forms long and polydisperse linear polymers (Guarné *et al.*, 2002, 2005). However, we have previously shown that mutation of Thr18, Ile21 or Ala25 abrogates filament formation (Guarné *et al.*, 2005). SeqA-A25R is a monodisperse dimer in solution, whereas SeqA-I21R and SeqA-T18E can form higher molecular-weight species at high concentrations (Guarné *et al.*, 2005). Hence, we selected SeqA-A25R (pAG8015) to elucidate the crystal structure of the SeqA dimer.

Residues 35–50 were disordered in the structure of the oligomerization domain of SeqA (Guarné *et al.*, 2005), as were residues 51–63 in the structure of the DNA-binding domain (Guarné *et al.*, 2002). Presumably, this region encompasses a flexible linker that provides plasticity to the SeqA–DNA interaction. In order to restrict the flexibility between the N- and C-terminal domains of SeqA-A25R, residues 41–59 were removed to generate the SeqA(Δ41–59)-A25R mutant (pAG8033). To this end, we designed self-complementary oligonucleotides that annealed with 15 nucleotides on each side of the deletion and used the QuikChange site-directed mutagenesis kit (Stratagene). The sequences of all mutants were verified by DNA sequencing (MOBIX Laboratory, McMaster University).

2.2. Overproduction and purification of SeqA(Δ41–59)-A25R

SeqA(Δ41–59)-A25R was overproduced in *E. coli* BL21(DE3) cells (Invitrogen) transformed with the pAG8033 plasmid. Cells were grown at 310 K to an OD_{600} of 0.7 and protein production was induced by the addition of 0.5 m*M* IPTG (isopropyl *β*-D-1-thio-galactopyranoside) to the culture. Cells were harvested after 3 h and washed with phosphate-buffered saline. Cell lysis was performed in buffer *A* (20 m*M* Tris pH 8, 100 m*M* NaCl, 5 m*M* DTT, 0.5 m*M* EDTA and 5% glycerol) with lysozyme (0.5 mg ml^{-1}) and brief sonication. Lysates were clarified by centrifugation (39 000*g* at 277 K for 40 min). The supernatant was loaded onto a heparin column (GE Healthcare) equilibrated with buffer *A* and SeqA(Δ41–59)-A25R was eluted from the column using a linear gradient to 1 *M* NaCl. The sample was further purified by ion exchange on a MonoS 10/100 GL column using the same buffers and gradient as for the heparin column. Purified SeqA(Δ41–59)-A25R was then concentrated to 3 mg ml^{-1} and stored in 20 m*M* Tris pH 8, 150 m*M* NaCl, 5 m*M* DTT, 0.5 m*M* EDTA and 5% glycerol. As expected from previous data (Guarné *et al.*, 2005), SeqA(Δ41–59)-A25R eluted at a volume consistent with a dimer from a size-exclusion column (Superdex 75 10/300 GL, GE Healthcare). All chromatographic steps were performed using an ÄKTA FPLC (GE Healthcare).

2.3. Complex formation and crystallization

Both methylated and unmethylated oligonucleotides were purchased from W. M. Keck Foundation at Yale University. Single-stranded oligonucleotides were purified over 10% polyacrylamide gels and eluted from the gel in elution buffer (10 m*M* Tris pH 7.5, 200 m*M* NaCl, 1 m*M* EDTA) overnight at 310 K. The eluted DNA was precipitated twice with 80 m*M* sodium acetate pH 7 and 70% ethanol at 253 K. Purified oligonucleotides were resuspended in deionized water and annealed to yield hemimethylated DNA duplexes as indicated in Table 1. SeqA(Δ41–59)-A25–DNA complexes (1:1 ratio) were incubated at room temperature for 15 min and subsequently stored at 277 K. Complex formation was monitored by size-exclusion chromatography (Superdex 200 10/300 GL, GE Healthcare) and electrophoretic mobility-shift assays.

Protein–DNA cocrystals were grown in hanging drops using the vapour-diffusion method at 277 K. Screening for crystallization conditions was performed using the Index (Hampton Research), Wizard I and II (Emerald) and Classics (Qiagen Inc.) crystallization screens. Initial hits were optimized using the sparse-matrix approach (Fig. 1). Diffraction-quality crystals were cryoprotected by addition

of 20% glycerol or increasing the amount of PEG 400 present in the crystallization drop where applicable.

2.4. Optimization of target DNA sequences

SeqA requires at least two hemimethylated GATC sequences residing on the same face of the DNA helix to form a stable complex (Brendler & Austin, 1999). The maximum separation between adjacent hemimethylated GATC sites that allows SeqA(Δ41–59)-A25R binding is 12 base pairs. However, complexes with GATC sequences at this spacing did not yield crystals (28/12, duplex length/GATC spacing) or the crystals diffracted to a very poor resolution (24/12, ∼20 Å).

In subsequent crystallization trials, the spacing between adjacent hemimethylated sites was limited to either nine or ten base pairs (bp), which exhibited the strongest binding (Guarné *et al.*, 2005). Additionally, the overall length was kept at about two helical turns (20–24 bp) to facilitate DNA packing within the crystal. Multiple combinations of these two parameters were screened using blunt-ended duplexes (Table 1; duplexes A–C and data not shown). All these hemimethylated duplexes formed stable complexes with SeqA(Δ41–59)-A25R and rendered nicely shaped crystals that diffracted X-rays to low resolution (Table 1 and Fig. 1). To facilitate end-to-end stacking of DNA molecules, we designed a 21/9 duplex with one overhanging nucleotide on each end (Table 1; duplex D). Use of duplex D led to the formation of better quality crystals as judged by their ability to diffract X-rays and hence we maintained the overall parameters of this duplex in subsequent optimizations.

Since SeqA interacts with the hemimethylated GATC site through the major groove of the duplex (Guarné *et al.*, 2002), a complex between the SeqA(Δ41–59)-A25R dimer and any GATC pairs separated by 9–10 bp would leave one face of the DNA duplex completely exposed to the solvent. In such a situation, DNA–DNA contacts could contribute significantly to the crystal packing. Therefore, we modified duplex D to engineer a CG snap-fastener on the opposite face of the GATC sites (duplex E in Table 1). CG dinucleotides have been previously shown to act as packing-driving boxes for oligonucleotides of different sizes (Timsit & Moras, 1994). However, this modification did not affect the crystallization conditions nor improve the diffraction quality of the crystals (Table 1), suggesting that the central part of the duplex did not mediate packing contacts.

Lastly, we optimized the ends of the duplexes to favour end-to-end interactions between neighbouring DNA duplexes. To this end, identical duplexes with either one or two overhangs at each end were

Table 2
Data-collection statistics for SeqA(Δ41–59)-A25R–duplex F cocrystals.

Values in parentheses are for the highest resolution shell.

Experimental conditions	
X-ray source	X29 (NSLS, BNL)
Wavelength (Å)	0.9795
Temperature (K)	100
Detector	ADSC Q315 CCD
No. of images	100
Exposure time (s)	5
Oscillation angle (°)	1
Data processing	
No. of measured reflections	293384
No. of unique reflections	30853
Space group	*R*3
Unit-cell parameters (Å, °)	$a = b = 121.8$, $c = 279.1$, $\gamma = 120$
Resolution (Å)	30.00–3.00 (3.11–3.00)
Completeness (%)	99.8 (100.0)
Multiplicity	3.2 (3.2)
Mean $I/\sigma(I)$	13.5 (1.8)
$R_{\rm merge}$	0.088 (0.527)
No. of complexes per ASU	2
Matthews coefficient $V_{\rm M}$ (Å^3 Da^{-1})	4.01
Solvent content (%)	72.6

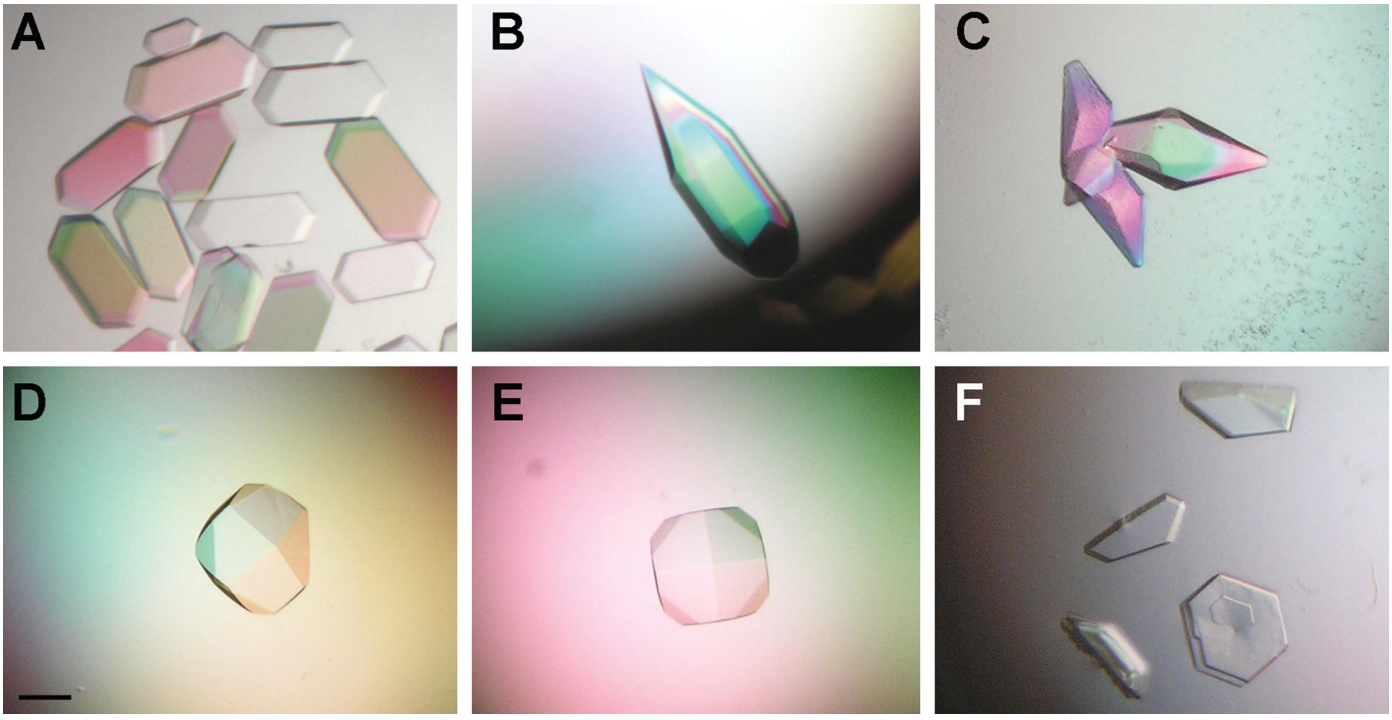

Figure 1
Crystals of SeqA(Δ41–59)-A25R bound to DNA. The duplex name is indicated in the top left corner of each image with the same convention as in Table 1. All pictures were taken at the same magnification; the scale bar in the bottom left image indicates 100 µm.

used (compare duplexes E and F in Table 1). Extending the single-stranded part of the duplex from one nucleotide to two nucleotides had a dramatic influence on crystal formation. Indeed, crystals of SeqA(Δ41–59)-A25R bound to duplex F had a radically different morphology to all the other cocrystals (Fig. 1).

2.5. Data collection, diffraction analysis and structure determination

X-ray diffraction data were collected from flash-frozen crystals on the X29 beamline of the National Synchrotron Light Source (Brookhaven National Laboratory). All data sets were indexed, integrated and scaled using the *HKL*-2000 package (Otwinowski & Minor, 1997). Crystals of SeqA(Δ41–59)-A25R bound to duplex F diffracted X-rays to 3.0 Å resolution and belonged to space group *R*3, with unit-cell parameters $a = b = 121.8$, $c = 279.1$ Å, $\gamma = 120^\circ$ (Fig. 2 and Table 2).

The structure of this complex was determined by molecular replacement using the N- and C-terminal domains of SeqA (PDB codes 1xrx and 1lrr, respectively). DNA was omitted from the search model in order to validate the quality of the molecular-replacement solution. Four DNA-binding domains and one dimerization domain were readily placed using *Phaser* (McCoy *et al.*, 2007). The two DNA duplexes within the asymmetric unit were easily identified in the initial electron-density map, as was the second dimerization domain.

3. Discussion

Extensive oligonucleotide modification yielded SeqA(Δ41–59)-A25R–DNA crystals that diffracted X-rays to 3.0 Å resolution. Three modifications of the DNA sequence were key to obtaining these cocrystals: (i) setting the GATC spacing to nine base pairs, (ii) fixing the duplex length to 20 base pairs and (iii) the inclusion of di-nucleotide overhangs. As revealed from the molecular-replacement solution, all three modifications improved the packing of the SeqA(Δ41–59)-A25R–DNA complex. On the other hand, the CG snap-fastener did not have an effect on the diffraction quality of the crystals, probably because crystal packing was not mediated by groove–backbone interactions between neighbouring DNA molecules.

Unexpectedly, the overhanging base pairs did not mediate head-to-tail interactions between symmetry-related DNA molecules. Instead, they fold away from the duplex axes and interact with neighbouring protein molecules. The single-stranded dinucleotide on the methylated strand folds back and interacts with the C-terminal domain of the SeqA(Δ41–59)-A25R molecule bound to its 5′ hemimethylated GATC site. While the 5′ guanine (G1) is hydrogen bonded to the side chains of Lys136 and Glu125 (Fig. 3), adenine A2 does not interact with the protein, revealing why a single overhang could not support this interaction. We are currently assessing whether this interaction strengthens the protein–DNA complex or simply facilitates crystal packing. Conversely, the other end of the DNA duplex does not mediate such intimate contacts with the protein. The single-stranded

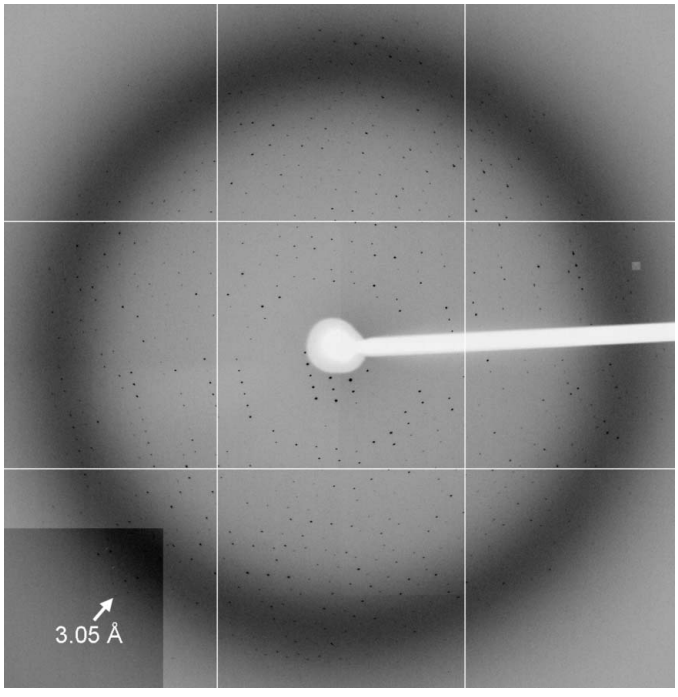

Figure 2
Representative X-ray diffraction pattern of SeqA(Δ41–59)-A25R bound to duplex F collected on beamline X29 (NSLS, BNL). Data were collected at a crystal-to-detector distance of 430 mm and a wavelength of 0.9795 Å. The resolution at the detector edge is 2.9 Å. In the bottom left inset the contrast has been adjusted to make weak reflections more prominent. One reflection at 3.05 Å resolution is indicated with a white arrow for reference.

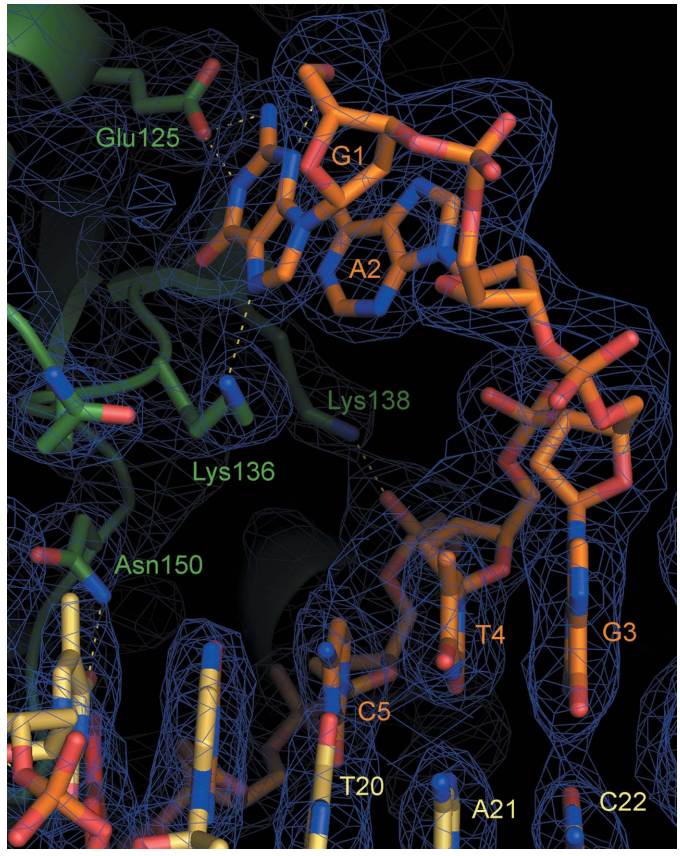

Figure 3
$2F_o - F_c$ electron-density map superimposed on the SeqA(Δ41–59)-A25R–DNA model at an early stage of refinement, showing the interactions between the two overhanging base pairs on the methylated strand and SeqA(Δ41–59)-A25R. The protein is shown as a green cartoon with interacting side chains highlighted as colour-coded sticks (C, green; N, blue; O, red). The sequence-specific interaction between Asn150 and T18 on the methylated A7–T18 base pair is also depicted. The unmethylated and methylated DNA strands are shown as pale yellow and orange sticks, respectively, with O and N atoms shown in red and blue for clarity. Hydrogen bonds are depicted as dashed yellow lines and the electron-density map (contoured at 1σ) as a blue mesh. This figure was prepared using *PyMOL* (DeLano, 2002).

dinucleotide on the unmethylated strand interacts with the N-terminal domain of a symmetry-related SeqA(Δ41–59)-A25R molecule. However, both T1 and C2 appear to be highly flexible as revealed by weak and disconnected electron density. In conclusion, the presence of two-nucleotide overhangs was the most important feature to obtain diffraction-quality crystals of SeqA(Δ41–59)-A25R bound to hemimethylated DNA. Our results indicate that the length of single-stranded overhangs may be a powerful variable to consider in protein–DNA cocrystallization even when end-to-end stacking of neighbouring DNA molecules is not expected.

We thank the PXRR staff at the NSLS (Brookhaven National Laboratory) for assistance during data collection. We are grateful to Monica Pillon for help with DNA purification. This work was supported by the Canadian Institutes of Health Research (MOP 67189).

References

Bach, T. & Skarstad, K. (2004). *Mol. Microbiol.* **51**, 1589–1600.
Barras, F. & Marinus, M. G. (1989). *Trends Genet.* **5**, 139–143.
Boye, E., Stokke, T., Kleckner, N. & Skarstad, K. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 12206–12211.
Brendler, T. & Austin, S. (1999). *EMBO J.* **18**, 2304–2310.
Campbell, J. L. & Kleckner, N. (1990). *Cell*, **62**, 967–979.
DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.*http://www.pymol.org.
Guarné, A., Brendler, T., Zhao, Q., Ghirlando, R., Austin, S. & Yang, W. (2005). *EMBO J.* **24**, 1502–1511.
Guarné, A., Zhao, Q., Ghirlando, R. & Yang, W. (2002). *Nature Struct. Biol.* **9**, 839–843.
Hiraga, S., Ichinose, C., Onogi, T., Niki, H. & Yamazoe, M. (2000). *Genes Cells*, **5**, 327–341.
Lu, M., Campbell, J. L., Boye, E. & Kleckner, N. (1994). *Cell*, **77**, 413–426.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Modrich, P. & Lahue, R. (1996). *Annu. Rev. Biochem.* **65**, 101–133.
Nievera, C., Torgue, J. J., Grimwade, J. E. & Leonard, A. C. (2006). *Mol. Cell*, **24**, 581–592.
Niki, H. & Hiraga, S. (1998). *Genes Dev.* **12**, 1036–1045.
Odsbu, I., Klungsoyr, H. K., Fossum, S. & Skarstad, K. (2005). *Genes Cells*, **10**, 1039–1049.
Onogi, T., Niki, H., Yamazoe, M. & Hiraga, S. (1999). *Mol. Microbiol.* **31**, 1775–1782.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Timsit, Y. & Moras, D. (1994). *EMBO J.* **13**, 2737–2746.