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

Structure of the C-terminal effector-binding domain of AhrC bound to its corepressor L-arginine

James A. Garnett,‡ Simon Baumberg,† Peter G. Stockley and Simon E. V. Phillips*

Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, England

‡ Current address: Division of Molecular Biosciences, Imperial College London, London SW7 2AZ, England.
† Deceased 11th April 2007.

Correspondence e-mail: s.e.v.phillips@leeds.ac.uk

Received 29 August 2007
Accepted 9 October 2007

PDB Reference: holo-CAhrC, 2p5m, r2p5msf.

The arginine repressor/activator protein (AhrC) from *Bacillus subtilis* belongs to a large family of multifunctional transcription factors that are involved in the regulation of bacterial arginine metabolism. AhrC interacts with operator sites in the promoters of arginine biosynthetic and catabolic operons, acting as a transcriptional repressor at biosynthetic sites and an activator of transcription at catabolic sites. AhrC is a hexamer of identical subunits, each having two domains. The C-terminal domains form the core of the protein and are involved in oligomerization and L-arginine binding. The N-terminal domains lie on the outside of the compact core and play a role in binding to 18 bp DNA operators called ARG boxes. The C-terminal domain of AhrC has been expressed, purified and characterized, and also crystallized as a hexamer with the bound corepressor L-arginine. Here, the crystal structure refined to 1.95 Å is presented.

1. Introduction

Bacillus subtilis synthesizes arginine from glutamine in a sequence of seven steps utilizing the gene products from *argB*, *argC*, *argD*, *argF*, *argG*, *argH* and *argJ* (Cunin *et al.*, 1986; Glansdorff, 1987). Carbamoyl phosphate, which is required for the synthesis of citrulline from ornithine, is the product of an additional enzyme, carbamoyl phosphate synthetase, encoded by the gene *cpa*. The bacteria are able to use arginine as a source of both carbon and nitrogen by catabolism back to glutamate *via* the arginase pathway. This requires the enzymes arginase (*rocF*), ornithine aminotransferase (*rocD*) and pyrroline-5-carboxylate dehydrogenase (*rocA*).

Harwood & Baumberg (1977) used arginine hydroxamate to identify mutant *B. subtilis* strains which overproduce arginine in nutrient-limiting media. Subsequent analysis of these mutants led to the cloning of the *ahrC* gene (Mountain & Baumberg, 1980; Mountain *et al.*, 1984), which was then shown to be the homologue of *argR*, the repressor controlling the arginine-biosynthesis genes in *Escherichia coli* (North *et al.*, 1989). AhrC shares 27% amino-acid sequence identity with ArgR and size-exclusion chromatography and analytical ultracentrifugation showed that it formed noncovalent hexamers at physiological concentrations (Czaplewski *et al.*, 1992), as does the *E. coli* protein (Lim *et al.*, 1987). AhrC is very sensitive to proteolytic cleavage, producing two similar-sized fragments (Czaplewski *et al.*, 1992) and suggesting that it contains two domains. A role for the N-terminal domains in binding to 18 bp operator sites (ARG boxes) and for the C-terminal domains in oligomerization and corepressor binding was identified through mutational studies of *E. coli* ArgR (Tian & Maas, 1994). The structure of intact apo-AhrC (Dennis *et al.*, 2002) in the absence of the corepressor L-arginine shows that the C-terminal domains make up the core of the protein with strict 32 symmetry, whilst the DNA-binding domains lie as loosely associated pairs around the periphery.

Here, we report the crystal structure of a hexamer of C-terminal domains of AhrC (CAhrC) bound with the corepressor L-arginine. We also discuss the structural differences from intact apo-AhrC (Dennis *et al.*, 2002) and the allosteric effects of ligand binding.

Table 1
Summary of data collection and refinement.

Values in parentheses are for the outermost resolution shell.

Data collection	
Beamline	SRS 14.2
Wavelength (Å)	0.980
Space group	$P6_522$
Unit-cell parameters	a = 117.01, b = 117.01, c = 68.93, α = 90, β = 90, γ = 120
Resolution (Å)	24.11–1.95 (2.06–1.95)
Unique observations	20787
R_{sym}	0.077 (0.391)
$\langle I/\sigma(I)\rangle$	32.4 (6.6)
Completeness (%)	99.9 (99.9)
Redundancy	16.7 (12.3)
Refinement	
R_{work}/R_{free} (%)	18.3/22.4
No. of residues	240
No. of ligands	5 L-arginines
No. of water molecules	169
R.m.s. deviations from ideal†	
Angles (°)	0.011
Bonds (Å)	1.42
Ramachandran plot†	
Most favoured (%)	92.9
Allowed (%)	6.1
Disallowed‡ (%)	1.0

† Determined using *PROCHECK* (Laskowski *et al.*, 1993). ‡ Residue Ser91 was found to be in a disallowed region of the Ramachandran plot in all three independent subunits of holo-CAhrC; however, the electron density supports this geometry.

2. Materials and methods

2.1. Cloning and expression

The apo-AhrC crystal structure (Dennis *et al.*, 2002) was used to define the C-terminal domain of AhrC (residues 67–149; CAhrC). A construct encoding this domain was obtained by PCR from the *ahrC* gene with primers containing *Nde*I and *Bam*HI restriction sites, ligated into the vector pET22b and transformed into *E. coli* strain BL21 (DE3), creating an initiator Met mutation (Asp67Met). Cells grown at 310 K in LB media were induced for 3 h with isopropyl β-D-1-thiogalactopyranoside at an $OD_{600\,nm}$ of 0.6.

2.2. Protein purification and crystallization

Cells were lysed by sonication in 20 m*M* Tris–HCl pH 7.5, 10 m*M* $MgCl_2$, 50 m*M* NaCl, 10 m*M* DTT, 1 m*M* PMSF, 25 µ*M* TPCK and 1% 2-propanol in the presence of 6 µg ml^{-1} lysozyme. DNaseI was added to the supernatant at 1 µg ml^{-1} and a 20%(*w*/*v*) ammonium sulfate precipitation was performed. The supernatant was dialysed against lysis buffer, loaded onto an SP Sepharose High Performance XK 16 column (GE Healthcare), equilibrated and washed with the same buffer. CAhrC eluted between 0.35 and 0.65 *M* NaCl and was dialysed overnight against 20 m*M* Tris–HCl pH 7.5, 200 m*M* NaCl, 10 m*M* L-arginine, 10 m*M* DTT, 1 m*M* PMSF, 25 µ*M* TPCK and 1% 2-propanol. It was then injected onto a HiLoad Superdex 75 XK 26/60 column (GE Healthcare), where holo-CAhrC eluted as a single 56 kDa peak with a single 9.3 kDa band on SDS–PAGE.

Crystallization experiments were carried out by hanging-drop vapour diffusion at 277 K. 2 µl of CAhrC (7 mg ml^{-1}) in 20 m*M* Tris–HCl pH 7.5, 100 m*M* NaCl, 10 m*M* L-arginine, 10 m*M* DTT, 1 m*M* PMSF, 25 µ*M* TPCK and 1% 2-propanol was mixed with an equal volume of well solution containing 100 m*M* Tris–HCl pH 8.8, 200 m*M* $MgCl_2$, 28%(*w*/*v*) PEG 400. Crystals grew after ∼14 d.

2.3. Data collection, structure determination and refinement

Crystals were frozen and data were collected on station 14.2 at Daresbury Synchrotron Radiation Source (SRS). All data were processed using *MOSFLM* (Leslie, 1992) and scaled using *SCALA* (Evans, 1993). Phases were obtained by molecular replacement using *AMoRe* (Navaza, 1994). Chains *A*, *B* and *C* from the *B. stearothermophilus* holo-ArgR core (Ni *et al.*, 1999; PDB code 1b4b) were used as the trial model and after rigid-body refinement the final solution resulted in a trimer with R = 0.44% and CC = 55.8%.

Approximately 10% of the data were used to calculate R_{free}. The initial model was refined in *CNS* (Brünger *et al.*, 1998) and model building was performed in *O* (Jones *et al.*, 1991). During the final stages of refinement, the holo-CAhrC coordinates (chains *A*–*C*) were submitted to the TLS motion-determination server (Painter & Merritt, 2006*a*,*b*). The modified PDB and TLS input files (with a single TLS parameter describing each chain) were used with *REFMAC*5 (Murshudov *et al.*, 1999) for a restrained and TLS *B*-factor refinement, and model building was performed with *Coot* (Emsley & Cowtan, 2004). The model shows good geometry as evaluated by *PROCHECK* (Laskowski *et al.*, 1993). All residues are in the allowed region of the Ramachandran plot except for Ser91; however, the electron-density map supports this geometry and it is an outlier in the apo-AhrC crystal structure (Dennis *et al.*, 2002). The data-collection and refinement statistics are shown in Table 1. Figures were generated with *PyMOL* (DeLano, 2002).

3. Results and discussion

3.1. Characterization

Purified CAhrC was analyzed by mass spectrometry and shown to have a molecular weight of 9280 Da (expected molecular weight 9280 Da; Garnett, 2005). Sedimentation-velocity experiments showed that at CAhrC concentrations of 100 µ*M* in the absence of L-arginine, ∼90% of CAhrC has a solution molecular weight of 51 230 Da, which is consistent with a hexamer (expected molecular weight of ∼55.7 kDa) in equilibrium with a trimer (molecular weight of ∼27.8 kDa). This compares with CAhrC in the presence of 10 m*M* L-arginine, which has a solution molecular weight of 52 994 Da under the same conditions and is ∼95% hexamer (Garnett, 2005).

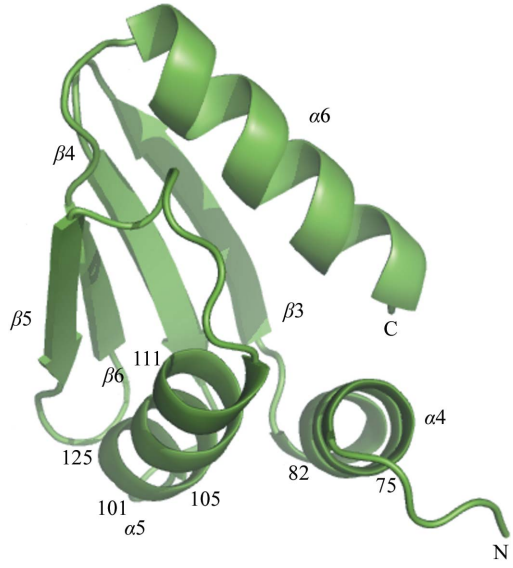

Figure 1
Secondary structure of the holo-CAhrC monomer, with some of the residues involved in the intra-trimer interface (Arg75, Asp82, Gly101 and Ala105) and L-arginine binding (Asp111 and Asp125) numbered.

3.2. Overall structure

The asymmetric unit contains one trimer, which was rebuilt from the molecular-replacement model and refined without NCS averaging to R = 18.3% and $R_{\rm free}$ = 22.4% at 1.95 Å resolution. The CAhrC monomer has α/β topology, comprising three α-helices and four β-strands (Fig. 1). Electron density is well defined and all residues except for some in the flexible linker (Met67*A*, Met67*B*–Phe70*B* and Met67*C*–Phe70*C*) could be modelled in the final structure.

The three independent monomers in the asymmetric unit are related by threefold NCS with average r.m.s.d.s for C^{α} atoms for residues 71–149 of 0.9 Å (chains *A* and *B*), 0.8 Å (chains *A* and *C*) and 0.4 Å (chains *B* and *C*). The trimer is mainly stabilized through interactions between three of the β-strands (β4, β5 and β6) in each monomer. On the external face of the trimer there is a shallow pore surrounded by hydrophilic residues around the threefold NCS axis, whilst on the buried face the binding of L-arginine creates a hydrogen-bond network linking Gln104, Asp111, Thr121, Cys123, Asp125, Asp126 and Thr127.

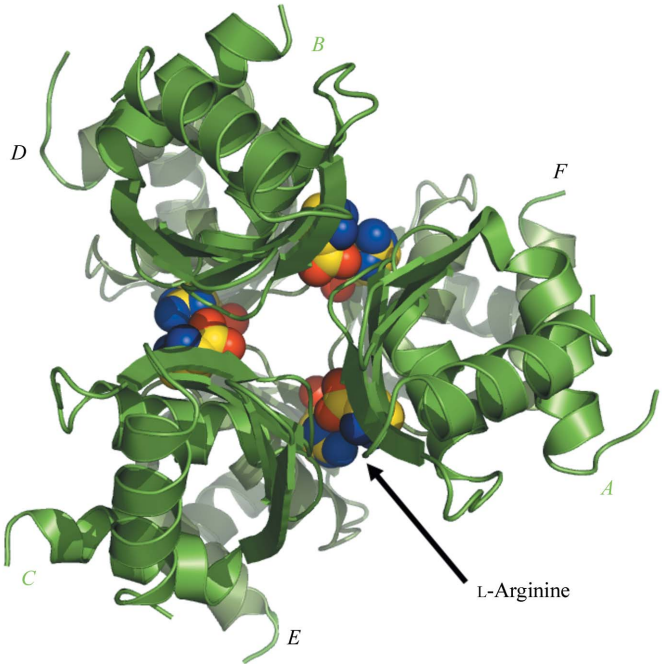

Figure 2
Secondary structure of the holo-CAhrC hexamer. The upper trimer is coloured light green, the lower trimer is dark green and L-arginine is shown as spheres. The crystallographic *c* axis relates chains *A* to *D*, *B* to *F* and *C* to *E*.

The other trimer that completes the holo-CAhrC hexamer is generated by a twofold rotation along the crystallographic *c* axis, which relates chains *A* to *D*, *B* to *F* and *C* to *E* (Fig. 2). The trimer–trimer interface is mediated by nonpolar residues that are mainly organized into buried patches and a series of direct and water-mediated polar contacts contributed by Arg75, Arg78, Asp82, Pro100, Gly101, Asn102, Ala105 and Asn112.

3.3. The L-arginine-binding pocket

Electron density for three L-arginine molecules, each bound between two subunits of one trimer and to a third subunit of the other trimer, was located adjacent to the conserved Asp125 (Fig. 3). Six L-arginine molecules bind within the interfaces of the CAhrC hexamer with a total of nine hydrogen bonds each, involving six residues (Gln104, Asp111, Cys123, Asp125, Asp126 and Thr127). Two of the hydrogen bonds to each L-arginine come from Asp125 of the opposing trimer, while the other seven are within the trimer carrying the binding pocket. An adjacent Asp125 also contributes to the pocket. In addition to the expected L-arginine molecules within the hexamer interfaces, two more were found in the electron-density map bound to the surface of the trimer within the solvent channel of the crystal, although only their guanidinium groups could be modelled with certainty.

3.4. Effects of L-arginine binding

There are minimal deviations between the intra-trimer interactions of apo-AhrC (Dennis *et al.*, 2002) and holo-CAhrC, with an average r.m.s.d. of 0.65 Å for all C^{α} atoms, so that each trimer behaves essentially as a rigid body. The main deviations involve the linker region (residues 68–70) that precedes α4. In apo-AhrC, the C-terminal domains interact, albeit loosely, with the DNA-binding domains (Dennis *et al.*, 2002) and this reduces the mobility of the linker. In holo-CAhrC, however, because the DNA-binding domains are absent these interactions are no longer possible and the structure of the linker is influenced by crystal contacts. However, there are quite significant changes between inter-trimer interactions when comparing apo-AhrC and holo-CAhrC. On L-arginine binding there is a rotation of one trimer with respect to the other by ∼15° (Fig. 4). In the apo-hexamer, the L-arginine-binding sites are not positioned to allow the corepressor guanidinium groups to interact with the exposed carboxylate groups of the Asp125 residues on the opposing trimer. The trimer rotation brings these groups into register, so that the interactions shown in Fig. 3 can form.

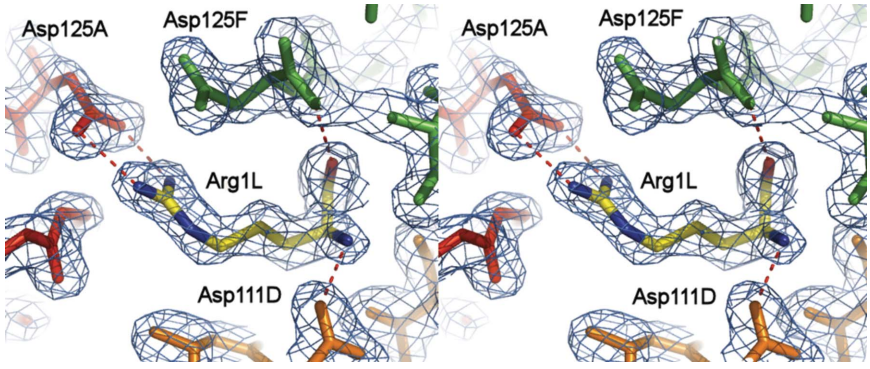

Figure 3
Stereoview of the L-arginine-binding pocket. The ligand is contacted by three different CAhrC subunits. The $2|F_o| - |F_c|$ map is contoured at 1.0 r.m.s., with chain *A* in red, chain *F* in green and chain *D* in orange and hydrogen bonds as red dashed lines. Two conserved Asp125 residues are shown, with the other ligand (ArgG1) just out of the plane of view.

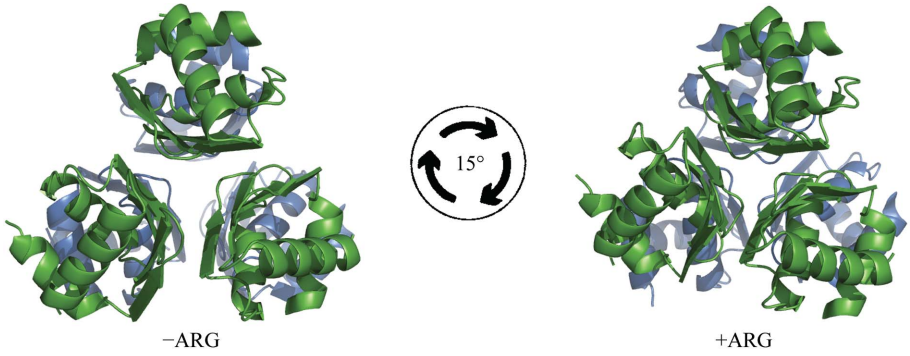

Figure 4
Conformational changes within the C-terminal domains of AhrC upon ligand binding. The C-terminal domain hexamer of apo-AhrC (residues 71–149) with the DNA-binding domains and linker removed (Dennis *et al.*, 2002) is shown on the left, while on the right is a hexamer of holo-CAhrC (residues 71–149) with the six L-arginine molecules not shown. The blue trimer is shown in the same orientation in apo-AhrC and holo-CAhrC, while the green trimer is rotated clockwise by ∼15° in response to L-arginine binding.

4. Conclusions

Isothermal titration calorimetry, surface plasmon resonance and proteolysis experiments with *E. coli* ArgR have shown that all six L-arginine-binding sites have equal intrinsic affinities for their ligand; however, occupancy of a single site causes a conformation change which resets the remaining five sites, making them 100-fold weaker (Jin *et al.*, 2004). However, in the crystal structures of *E. coli* apo- and holo-ArgR C-terminal domain hexamers no net reorientation of the trimers was observed (Van Duyne *et al.*, 1996). Only minor changes were observed in the N-terminal helices of the *E. coli* ArgR core, which were suggested would be propagated to the DNA-binding domains and be part of the mechanism of allosteric regulation.

Both of the *E. coli* crystal forms could be grown from the same conditions in the presence of an unusually wide range of salts and ionic strengths and over a pH range from 5 to 8 (Van Duyne *et al.*, 1996), suggesting that this crystal packing is especially favourable for this structure. If the apo-ArgR core is in equilibrium between different quaternary states, such as different trimer rotations, then the favourable crystal packing might displace this away from the apo-form towards the holo state, even if the latter is rare in solution. In addition, there is no structure of an intact *E. coli* ArgR hexamer and there are likely to be interactions between the N- and C-terminal domains which might affect the quaternary structure. This represents an alternative hypothesis for allosteric regulation of ArgR to that proposed (Van Duyne *et al.*, 1996), but would be consistent with the behaviour of the two *Bacillus* homologues. Rotation of the CAhrC trimers has also been observed in the crystal structure of the C-terminal domain holo-hexamer from *B. stearothermophilus* ArgR compared with its intact apo crystal structure (Ni *et al.*, 1999). In this case, as for AhrC, the rotation is most likely not a consequence of the crystal-packing environment, but is rather driven by the need to align the corepressor guanidinium groups with Asp125 and is presumably the origin of the allosteric regulation. The change would allow pairs of DNA-binding domains to move closer together, facilitating the formation of a dimer of N-terminal domains at the periphery of the repressor and presenting an active DNA-binding face to DNA containing ARG-box operator sites.

We thank the SRS for providing synchrotron-radiation facilities and the beamline scientists at station 14.2 for their help with data collection. We thank Dr M. Parsons for correspondence, Dr A. Ashcroft for performing mass spectrometry and Dr A. Baron for performing analytical ultracentrifugation experiments. This work was funded by the BBSRC.

References

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Cunin, R., Glansdorff, N., Pierard, A. & Stalon, V. (1986). *Microbiol. Rev.* **50**, 314–352.
Czaplewski, L. G., North, A. K., Smith, M. C. M., Baumberg, S. & Stockley, P. G. (1992). *Mol. Microbiol.* **6**, 267–275.
DeLano, W. L. (2002). *The PyMol User's Manual.* DeLano Scientific, San Carlos, California, USA.
Dennis, C. A., Glykos, N. M., Parsons, M. R. & Phillips, S. E. V. (2002). *Acta Cryst.* D**58**, 421–430.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, P. R. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp 114–122. Warrington: Daresbury Laboratory.
Garnett, J. A. (2005). PhD thesis. Department of Biochemistry and Molecular Biology, University of Leeds, England.
Glansdorff, N. (1987). *Escherichia coli and Salmonella: Cellular and Molecular Biology*, edited by F. C. Neidhardt, J. L. Ingraham, K. Brooks Low, B. Magasanik, M. Schaechter & H. E. Umbarger. Washington DC: American Society for Microbiology.
Harwood, C. R. & Baumberg, S. (1977). *J. Gen. Microbiol.* **100**, 177–188.
Jin, L., Xue, W.-F., Fukayama, J. W., Yetter, J., Pickering, M. & Carey, J. (2004). *J. Mol. Biol.* **346**, 43–56.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Lim, D., Oppenheim, J. D., Eckhardt, T. & Maas, W. K. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 6697–6701.
Mountain, A. & Baumberg, S. (1980). *Mol. Gen. Genet.* **178**, 691–701.
Mountain, A., Mann, N. H., Manton, R. N. & Baumberg, S. (1984). *Mol. Gen. Genet.* **197**, 82–89.
Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Ni, J., Sakanyan, V., Charlier, D., Glansdorff, N. & Van Duyne, G. D. (1999). *Nature Struct. Biol.* **6**, 427–432.
North, A. K., Smith, M. C. M. & Baumberg, S. (1989). *Gene*, **80**, 29–38.
Painter, J. & Merritt, E. A. (2006*a*). *Acta Cryst.* D**62**, 439–450.
Painter, J. & Merritt, E. A. (2006*b*). *J. Appl. Cryst.* **39**, 109–111.
Tian, G. & Maas, W. K. (1994). *Mol. Microbiol.* **4**, 599–608.
Van Duyne, G. D., Ghosh, G., Maas, W. K. & Sigler, P. B. (1996). *J. Mol. Biol.* **256**, 377–391.