Structural insight into gene transcriptional regulation and effector binding by the Lrp/AsnC family

Paul Thaw, Svetlana E. Sedelnikova, Tatyana Muranova, Sebastian Wiese, Sylvia Ayora[1,2], Juan C. Alonso[2], Arie B. Brinkman[3], Jasper Akerboom[4], John van der Oost[4] and John B. Rafferty*

Department of Molecular Biology and Biotechnology, Krebs Institute, University of Sheffield, Western Bank, Sheffield S10 2TN, UK, [1]Departamento de Biologia Molecular, Universidad Autonoma de Madrid, Cantoblanco, 28049 Madrid, Spain, [2]Departamento de Biotecnologia Microbiana, Centro Nacional de Biotecnologia, CSIC, Campus Universidad Autonoma de Madrid, Cantoblanco, 28049 Madrid, Spain, [3]Department of Molecular Biology, NCMLS M850/3.79, Geert Grooteplein 30, 6525 GA, Nijmegen, The Netherlands and [4]Laboratory of Microbiology, Wageningen University, Hesselink van Suchtelenweg 4, 6307 CT Wageningen, The Netherlands

ABSTRACT

The Lrp/AsnC family of transcriptional regulatory proteins is found in both archaea and bacteria. Members of the family influence cellular metabolism in both a global (Lrp) and specific (AsnC) manner, often in response to exogenous amino acid effectors. In the present study we have determined both the first bacterial and the highest resolution structures for members of the family. *Escherichia coli* AsnC is a specific gene regulator whose activity is triggered by asparagine binding. *Bacillus subtilis* LrpC is a global regulator involved in chromosome condensation. Our AsnC-asparagine structure is the first for a regulator–effector complex and is revealed as an octameric disc. Key ligand recognition residues are identified together with a route for ligand access. The LrpC structure reveals a stable octamer supportive of a topological role in dynamic DNA packaging. The structures yield significant clues to the functionality of Lrp/AsnC-type regulators with respect to ligand binding and oligomerization states as well as to their role in specific and global DNA regulation.

INTRODUCTION

Proteins belonging to the Lrp/AsnC family of global or specific transcriptional regulators are widely distributed in numerous prokaryotes, including bacteria and archaea (1,2). At least one Lrp-like homologue can be identified in 45 and 94% of the currently available bacterial and archaeal genomes,

respectively. To date, there are no confirmed homologues in the available eukaryal genomes, indicating that the family is probably restricted to prokaryotes (3). Members of the Lrp/AsnC family typically have a molecular mass of ∼15 kDa but populate a range of multimeric species in solution, which include dimers, tetramers, octamers and hexadecamers (4–8). In addition to their well-studied role in specific and global regulation of gene expression, it has been suggested that some bacterial Lrp homologues may play a role in (i) chromosome structure and organization (9) based upon observations of high copy number (between 1300 and 3200 dimers per cell) (10), (ii) DNA bending (11) and (iii) condensation of DNA into globular nucleoprotein-like structures (6). Several homologues from Archaea have also been characterized recently. These studies concerned regulators that block one or more binding sites of the general transcription initiation machinery, and as such repress the expression of the downstream gene [e.g. *Pyrococcus furiosus* LrpA (7) and *Sulfolobus solfataricus* LrpB (12)]. An additional homologue from *S.solfataricus* (LysM) appears to be a lysine-dependent transcription activator (13). The best-documented case of an archaeal Lrp-like transcription activator is *Methanocaldococcus jannaschii* Ptr2 that enhances the expression of several genes in a ligand-independent manner (14,15). *S.solfataricus* Lrs14, a relatively abundant protein (2,16), may correspond to a chromosome organizer, as has been suggested for some bacterial Lrp homologues and for *Pyrococcus* OT3 FL11, where glutamine triggers the binding and wrapping of DNA (17).

The best-characterized member of the family is *Escherichia coli* Lrp, which controls a global regulon encompassing at least 10% of all *E.coli* genes (18). The genes that belong to the Lrp-regulon encode proteins that are involved in transport,

---

*To whom correspondence should be addressed. Tel: +44 (114) 222 2809; Fax: +44 (114) 222 2800; Email: j.rafferty@sheffield.ac.uk

degradation and biosynthesis of amino acids, as well as a small number of proteins involved in the production of pili, porins, sugar transporters and nucleotide transhydrogenases (19,20). *E.coli* Lrp utilizes the binding of L-leucine to trigger either activation or repression of some target promoters, although in the majority of cases control by Lrp is leucine-independent, e.g. negative autoregulation of its own *lrp* gene.

E.coli AsnC shows notable sequence identity (25%) with Lrp (Figure 1), which has resulted in them being classified as a distinct evolutionary protein family (21). In contrast to Lrp, AsnC has only been shown to exert specific control of its own gene and that of *asnA*. The latter gene codes for asparagine synthase, and is regulated by AsnC in an asparagine-dependent fashion. Increasing levels of exogenous L-asparagine reduces *asnA* transcription leading to decreased cellular levels of asparagine synthase, consistent with a classical negative feedback mechanism (22,23). AsnC has also been demonstrated to autoregulate its own expression in an asparagine-independent manner.

Bacillus subtilis LrpC was one of seven genes encoding proteins belonging to the Lrp/AsnC family identified in its genomic sequence. It shares 34 and 25% identity with *E.coli* Lrp and AsnC, respectively, and has reported functionality in both sporulation and amino acid metabolism (24). LrpC has been shown to bind multiple sites in the upstream region of its own gene, resulting in slight positive autoregulation, in contrast to the negative autoregulation observed for other family members (25). Gel filtration studies indicated that LrpC forms a tetramer in solution and DNA binding has been reported to proceed cooperatively in a sequence independent manner, with LrpC preferentially recognizing intrinsically curved regions of DNA (26). Interestingly, electron microscopy studies have demonstrated that LrpC is able to form nucleoprotein complexes capable of wrapping DNA into a right-handed super-helix to form structures resembling nucleosomes (27) Furthermore, LrpC has been demonstrated to constrain DNA supercoils, implying it may also have a role akin to bacterial chromatin (26).

Prior to this study, structures of family members existed only for the archaeal proteins, *Pyrococcus furiosus* LrpA (28) and *Pyrococcus* OT3 FL11 (17). These revealed an N-terminal DNA-binding domain, containing a helix–turn–helix (HtH)
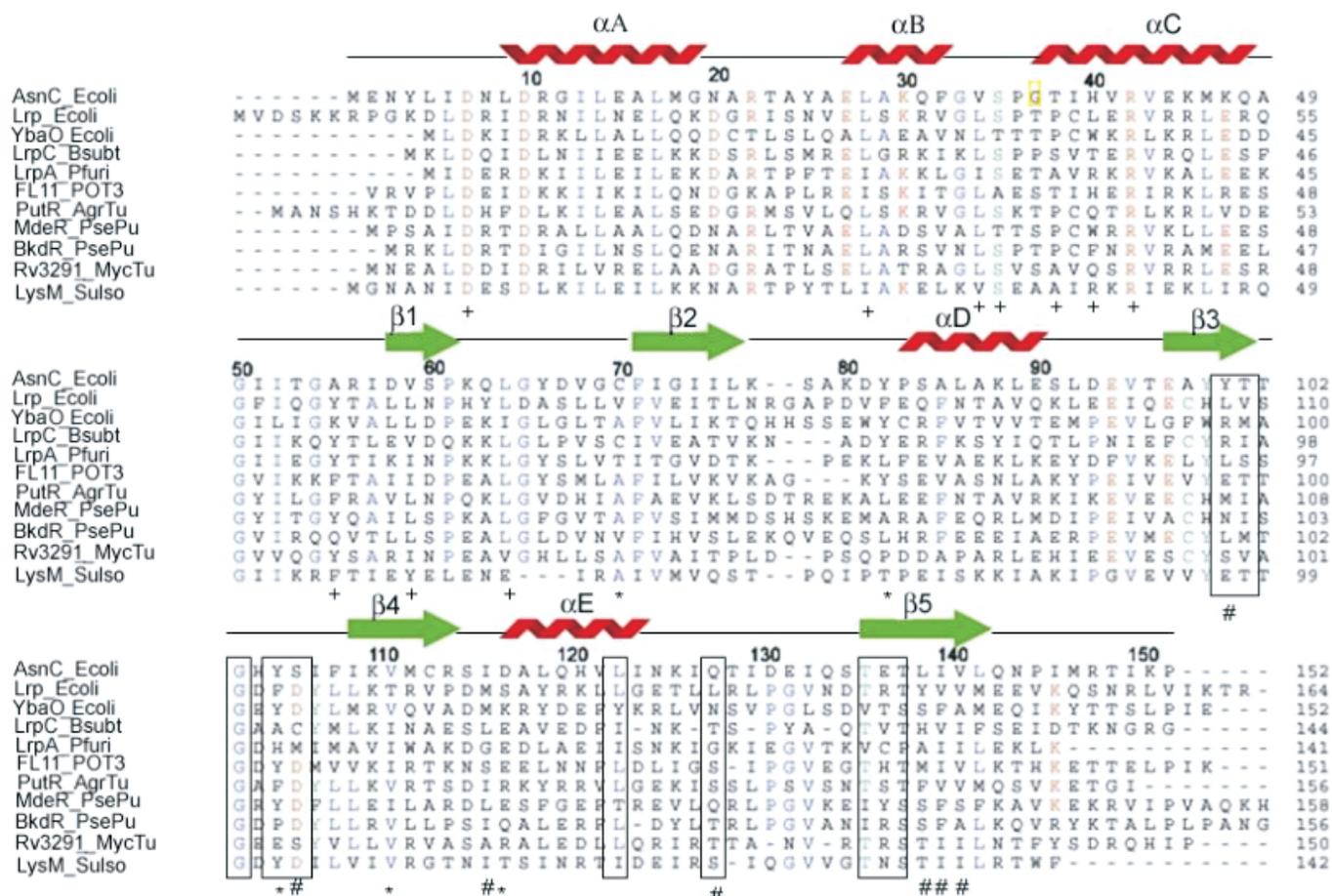
Figure 1. Sequence alignment of the Lrp/AsnC family. A structure-based multiple sequence alignment of the AsnC/Lrp family is shown and the residues forming the ligand-binding site are identified. Secondary structure elements are indicated as red for α-helices and green for β-strands. Residues are coloured as blue for hydrophobic, red for charged and green for polar. The location of the G37 to E37 mutation in our AsnC construct is marked with a yellow box. Residues found to be important in the formation of the ligand-binding site are indicated by closed boxes. The positions of DNA binding, activation and leucine response mutants identified in *E.coli* Lrp are indicated by the symbols plus, asterisk and hash, respectively. The five C-terminal residues of BkdR (MTLRE) have been omitted from the alignment. The figure was produced using the INDONESIA alignment package (D. Madsen, P. Johansson and G.J. Kleywegt manuscript in preparation). Species abbreviations are as follows: Ecoli, *E.coli*; Bsubt, *B.subtilis*; Pfuri, *P.furiosus*; POT3, *Pyrococcus OT3*; AgrTu, *A.tumefaciens*; PsePu, *P.putida*; MycTu, *Mycobacterium tuberculosis*; SulSo, *S.solfataricus*.

motif coupled via a linker region to a C-terminal effector domain. The latter domain was found to be reminiscent of the ACT family of small molecule binding domains; however, because of significant structural and functional differences it has been termed the RAM domain (regulation of amino acid metabolism) (29). The main structural unit of the protein consists of a homo-dimer, held together mainly by interactions between the anti-parallel β-sheets of the C-terminal domain. Analysis of sequence comparisons and mapping of mutational data (30) onto these structures revealed no clear ligand-binding site, only that it was most likely present in the interface between the dimers (28). On the basis of an alignment of RAM domains, certain conserved residues (equivalent to L95, M101, A134, I135 and I136 of *P.furiosus* LrpA) have been predicted to be involved in ligand binding (29).

We have determined the crystal structures of both *E.coli* AsnC and *B.subtilis* LrpC to 2.4 Å resolution. The AsnC structure is seen to be an octamer with the L-asparagine ligand bound in a cleft at the interface between dimers. Analysis of this structure with respect to biochemical and mutational data has implications for the oligomerization state of the protein *in vivo* and its subsequent DNA binding. The structure of LrpC is also revealed to be octameric and yields clues as to how nucleosome-like structures might be formed.

MATERIALS AND METHODS

A full account of the purification, crystallization and data collection details for both proteins will be published elsewhere (P. Thaw, S. Sedelnikova, S. Ayora, J. Van der Oost and J.B. Rafferty, manuscript in preparation). In brief, a glycine residue 37 to glutamate variant (G37E) of AsnC was over-expressed in the methionine auxotroph *E.coli* B834(DE3) strain using a pLUW634 vector (a derivative of pET24d carrying the *asnC* gene of *E.coli* in which the unintentional G37E mutation was introduced during PCR amplification). A two-step chromatographic purification was then applied, firstly utilizing ion-exchange chromatography on DEAE-Sepharose (Amersham Biotech), followed by gel filtration on a Superdex S-200 column (Amersham Biotech). Sample purity was assessed by SDS–PAGE, prior to concentrating the protein to 10 mg/ml for hanging drop vapour diffusion trials at 17°C. Initial crystals were obtained in 2.2 M (NH$_4$)$_2$SO$_4$ in 0.1 M Bicine, pH 8.7, plus 5 mM L-asparagine (Table 1). Crystal optimization resulted in square plate-like crystals up to dimensions of 0.3 × 0.3 × 0.05 mm in 2 days. Crystals were found to belong to space group P4 and contain two copies of the AsnC monomer in the asymmetric unit. Crystals of a selenomethionine-incorporated form of the protein were also grown at 14 mg/ml protein in 8% PEG 8K, 200 mM NH$_4$I, in 0.1 M MES, pH 5.6. These crystals were found to belong to the space group I222, with 10 copies of AsnC in the asymmetric unit. The selenium substructure of these latter crystals was determined using SHELXD (31) from a multi-wavelength anomalous dispersion (MAD) experiment (32) on station BM14 at the European Synchrotron Radiation Facility (ESRF). Data were processed using the MOSFLM package (33) and scaled in SCALA (34). An initial model was built in TURBO-FRODO (35) using maps and phases calculated in both SHARP/SOLOMON (36,37) and the CCP4 suite (38) to a

Table 1. Summary of data collection and refinement statistics

	AsnC Native 1	AsnC Se derivative (peak)	LrpC Se derivative (peak)
Data collection			
Space group	P4	I222	C222$_1$
Cell constants (Å)	$a = b = 103.1$	$a = 121.0$	$a = 165.8$
	$c = 52.4$	$b = 179.8$	$b = 166.1$
		$c = 201.6$	$c = 155.8$
Vm (Da Å^{-3})	4.1	3.2	4.1
Wavelength (Å)	1.488	0.97	0.97
Resolution (Å)	2.4 (2.53–2.4)	3.0 (3.2–3.0)	2.4(2.53–2.4)
Observations	76 065	352 929	329 639
Unique observations	21 843	24 268	83 371
R$_{merge}$[a]	9.8 (63.9)	10.4 (32.2)	9.3 (49.1)
Completeness (%)	99.2 (99.9)	99.6 (99.6)	99.4 (99.8)
$\langle I \rangle / \langle \sigma I \rangle$	9.8 (1.7)	7.9 (2.8)	9.6 (1.9)
Refinement			
R_{factor} (%)[b]	23		22
R_{free} (%)[c]	29		26
r.m.s.d.			
Bond lengths (Å)	0.01		0.01
Bond angles (°)	0.97		1.23
No. of residues	304		1120
No. of waters	142		240

Values in parentheses are for data in the highest resolution shell.
[a]$R_{merge} = \Sigma |I - \langle I \rangle| / \Sigma I$, where I is the integrated intensity of a given reflection.
[b]$R_{factor} = \Sigma |F_{obs} - F_{calc}| / \Sigma F_{obs}$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.
[c]R_{free} was calculated using a randomly selected 5% of the data excluded from all stages of the refinement.

resolution of 3.0 Å. The 10 copies of the protein in the asymmetric unit comprised an octamer and a dimer (which formed an additional octamer by the application of crystallographic symmetry). Crystallographic refinement of these data proved unsuccessful, largely due to significant anisotropy. Thus this model was used in the molecular replacement program MOL-REP (39) to solve a dataset of the other crystal form in space group P4 at 2.4 Å resolution obtained from station 14.1 at the SRS Daresbury.

B.subtilis LrpC was overexpressed in *E.coli* BL21(DE3) pLysS using a pET3b vector. Purification was achieved via an (NH$_4$)$_2$SO$_4$ cut at 2.2 M, followed by gel filtration on a Hi-Load Superdex 200 column. Samples were assessed for purity by SDS–PAGE and concentrated to 10 mg/ml for hanging drop crystallization trials. Crystal optimization resulted in large triangular shaped crystals (0.3 × 0.25 × 0.15 mm) in 2.2 M (NH$_4$)$_2$SO$_4$ 0.1 M Bicine pH 8.7. Crystals were found to belong to the space group C222$_1$ suggesting an octamer in the asymmetric unit. The structure was determined via MOLREP using an octameric model of *P.furiosus* LrpA in combination with selenomethionine derivative data collected at the ESRF. A set of initial phases was generated using the CCP4 suite, prior to solvent flattening and phase extension in DM (40) to produce a 2.4 Å map.

The final AsnC and LrpC models were built using the program COOT (41) and refined in REFMAC (42) with stereochemical monitoring carried out using the program PRO-CHECK (43). AsnC and LrpC were both refined at 2.4 Å to crystallographic R-factors of 20.6% ($R_{free} = 27.2\%$) and 22.9% ($R_{free} = 26.7\%$), respectively. The solvent structures for both AsnC and LrpC have been modelled using wARP (44) and contain 143 and 240 water molecules, respectively.

All figures were produced using PyMOL (http://www.pymol.org) with the exception of the electrostatic potential surfaces, which were produced in GRASP (45). The atomic coordinates and structure factors for AsnC and LrpC are deposited at the protein data bank with codes 2cg4 and 2cfx, respectively.

RESULTS

Overall structure

The monomer structure of both AsnC and LrpC comprises two domains. An N-terminal DNA-binding domain consisting of 3 α-helices (αA–αC), with αB–αC forming a simple 2 helix motif (HtH) (46) stabilized by a hydrophobic core protected by αA (Figure 2a–d). The G37 to E37 mutation in our AsnC construct (Materials and Methods) lies on the surface at the N-terminal end of helix αC. This represents a significant change in both size and charge and whilst its location suggests it might impair DNA binding, it should have no other detrimental effect to the overall fold of the protein. Crystallization trials with the wild-type protein produced only very poorly diffracting crystals whose space group could not be determined. The G37E variant yielded diffracting crystals suitable for crystallographic studies and analysis of the packing reveals E37 in some subunits does contribute to the crystal lattice contacts. The packing does not resemble the helical arrangement seen with *Pyrococcus* OT3 FL11 (17). The relative overall orientation of the N- and C-terminal domains is essentially the same in both AsnC and LrpC. However there is a shift of 20° in the position of helix αC relative to strand β1 between AsnC and LrpC (Figure 2). The inter-helix spacing between helices αB and αC within the N-terminal domain is maintained between the two structures by increased curvature of helix αC in LrpC. This DNA-binding domain is linked by a single β-strand (β1) to a C-terminal effector binding or RAM domain (29) formed by a four-stranded anti-parallel β-sheet (β2–β5) flanked on one face by two α-helices (αD–αE).

A dimer representing one of the functional units in AsnC or LrpC is stabilized via the formation of a hydrophobic core between strands β2–β5 of each monomer. The dimer is further stabilized via direct hydrogen bonding interactions between residues on β1 of each monomer, which come together to form a 2-stranded anti-parallel β-ribbon. Predominantly hydrophobic interactions between the C-terminus of one monomer and strand β3 and helix αD of another also contribute to dimer stability.

The P4 symmetry of the AsnC crystal lattice allows the formation of an octamer, consisting of a tetramer of dimers ∼52 × 52 × 118 Å in size (Figure 2e and g). The asymmetric unit of the LrpC crystal contains an octamer measuring 56 × 56 × 118 Å in size. The dimer–dimer interface consists mainly of hydrophobic contacts, with the main interactions forming between helix αE and strand β5 of one dimer pair, with residues leading into helix αD and those in a loop between strands β3 and β4 of another. Analysis using AREA-IMOL (47) with a search probe radius of 1.4 Å has provided measurements of buried surface area upon dimer and octamer formation (Table 2). A comparison of these values with those of other multimeric proteins (48) suggests that these values fall

within the normal range for cytosolic proteins from mesophilic sources.

Asparagine binding in AsnC

In the refined structure of the AsnC octamer, clear additional density was observed in the cleft between the turn from strand β3 to β4 of one monomer and strand β5 of another. This additional density was identified as bound L-asparagine, which had been present at a concentration of 5 mM during crystal growth. Eight molecules of asparagine were identified in each octamer. Loop residues Y100 to S106 from one dimer pair form one side of a binding pocket, with the side chain of the Y100 pointing inwards together with the backbone carbonyls of T101 and G103 (Figure 3a). Residue G103 adopts a positive phi angle of ∼90° to create a sharp turn that allows the use of the backbone carbonyls of T101 and G103 in binding the effector. The remainder of the pocket is constructed from residues contributed by a second dimer pair, Q128 and L123 emanating from helix αE, and residues T136 to T138 of strand β5 (Figure 3b). The asparagine is positioned such that the generic carboxyl and amino groups of the amino acid display hydrogen bonding potential to the backbone amides or carbonyls of G103, T138 and Y105. The side chain carboxyamide of the asparagine is most likely stabilized through interactions with the side chains of Y100, Q128 and the backbone of T101 and S106. AsnC discriminates between asparagine and aspartate and an analysis of the binding pocket suggests that unfavourable interactions would arise between the charged sidechain of the effector molecule and residue E128. The formation of a stable octamer in the presence or absence of asparagine is observed by gel filtration (data not shown) suggesting that the ligand is not an absolute requirement for octamer formation, which is consistent with the previously obtained effector-free structure of *P.furiosus* LrpA (28), as well as with reported observations of effector-free octamers in studies on *E.coli* Lrp (49). An analysis of the molecular surfaces of the octameric state of AsnC using the program GRASP (45) reveals channels that allow the diffusion of asparagine into the protein (Figure 4). Residues H104 to S106 and I75 to L76 of one dimer pair, together with a contribution from S135 to E137 from an adjacent dimer, shape the channel entrance. The eight channels branch off the central cavity of the octamer and would permit diffusion of the ligand in and out of the protein, allowing for a rapid response to exogenous asparagine levels.

DISCUSSION

Mutational and sequence analysis

Effector-binding region. The structural data available for four members of the family, together with our ligand-binding data allow us to study the likely effects of mutations both on effector binding and the oligomeric state of the protein. Previous work on *E.coli* Lrp (50) identified subgroups of residues which influenced either DNA binding or its interaction with leucine. In Lrp, variants found to be insensitive to leucine carried mutations in one of seven positions in the C-terminal domain (residues L107, D114, M124, L136, Y147, V148 and V149). In addition, a thorough comparative analysis of RAM
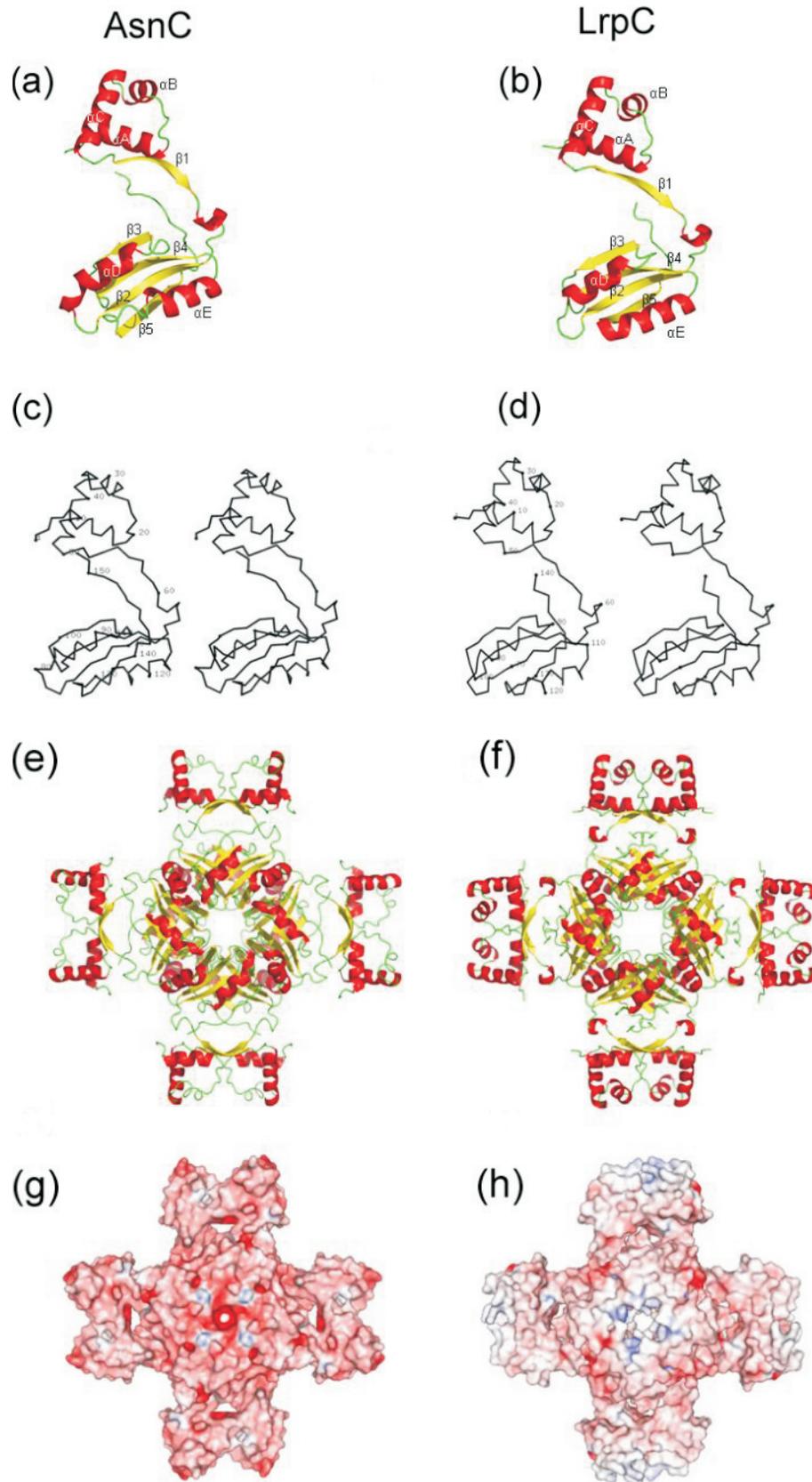
Figure 2. Overall fold of AsnC and LrpC proteins. Schematic representations of the monomer fold of *E.coli* AsnC (**a**) and *B.subtilis* LrpC (**b**). α-Helices are coloured red and β-strands yellow. Stereo representation of the Cα backbone of monomers of AsnC (**c**) and LrpC (**d**). Cartoon representations of the octameric forms of AsnC (**e**) and LrpC (**f**) together with their respective electrostatic potential surfaces (**g** and **h**). The surfaces are coloured by electrostatic potential where red is (⩽−30 kcal(mol·e)$^{-1}$) and blue (⩾30 kcal(mol·e)$^{-1}$).
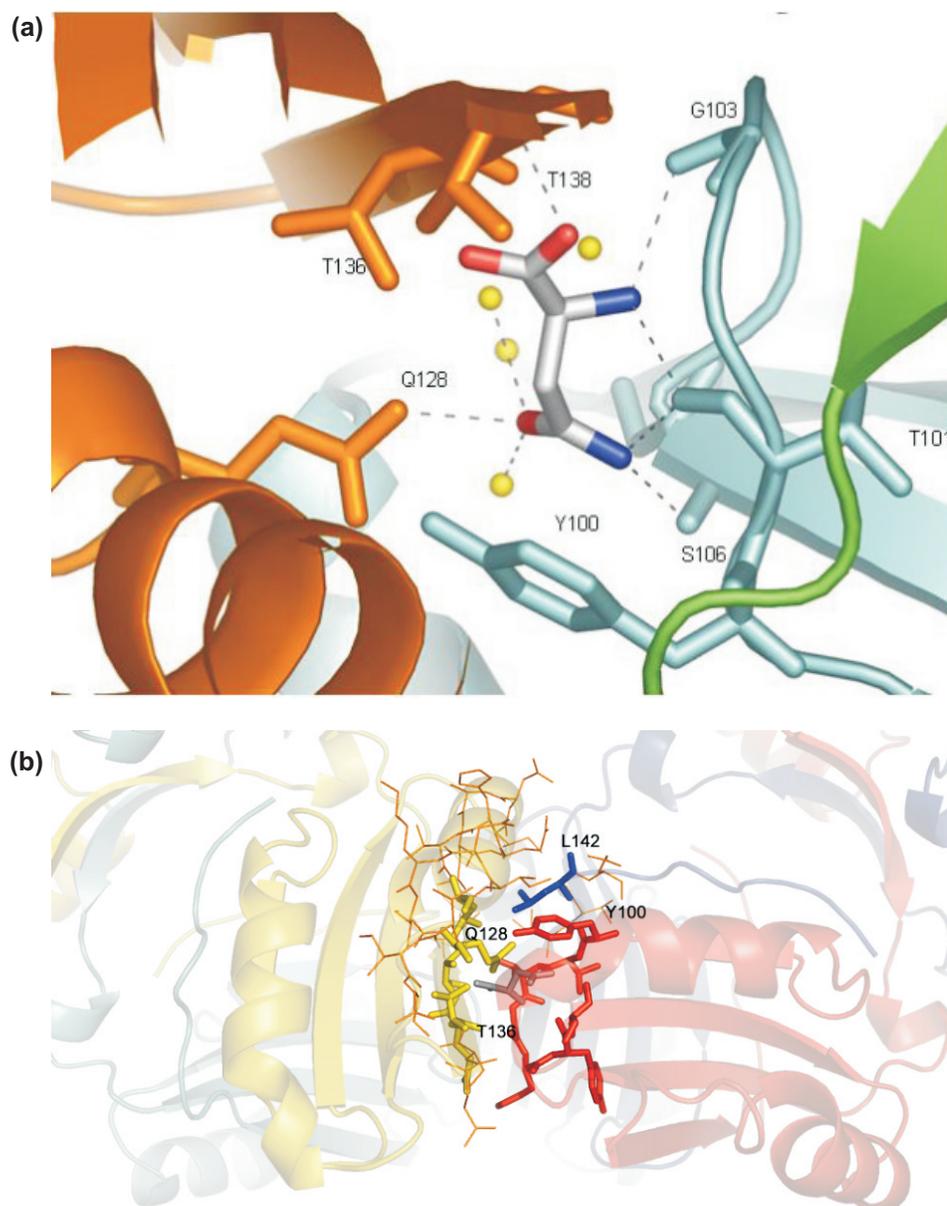
sequences resulted in the identification of conserved regions that agreed well with the previous mutagenesis work. Two potential ligand-binding sites were predicted: one between the loop connecting strand β3 to β4 of the monomer and

Table 2. Summary of surface area buried upon multimerization

	AsnC	LrpC
% and area of monomer buried upon dimer formation	29 (2954 Å²)	26 (2529 Å²)
% and area of monomer buried upon octamer formation	13 (1335 Å²)	9 (938 Å²)

the strand β5 of another, together with a second site the mirror image of this at the dimer–dimer interface (29).

The Lrp leucine response-mutations correspond to the following residues of AsnC: Y100, S106, I116, Q128, L139, I140 and V141 (Figure 1). The AsnC crystal structure reveals that three of these residues are directly involved in effector-binding pocket formation (Y100, S106 and Q128). Residues L139 to V141 are involved in hydrophobic packing interactions, which stabilize key areas surrounding the binding site. The position of residues T136, E137 and T138 (strand β5) of one monomer together with the critical β3 to β4 turn (of the second monomer forming the dimer) involved in ligand

Figure 3. Asparagine binding pocket of AsnC. (**a**) Stick representation of the asparagine-binding pocket in the AsnC octamer. Residues contributed by a monomer in the first dimer are shown in brown, with those from monomers in a second dimer pair coloured blue and green. Water molecules are indicated by yellow spheres. Hydrogen bonds are indicated as dashed lines. The asparagine is coloured with oxygen atoms in red, nitrogen atoms in blue and carbon in grey. (**b**) A representation of the Asn binding pocket at the dimer–dimer interface. The dimer to the right is displayed with monomers in red and blue transparent cartoon, with the dimer on the left shown with monomers in cyan and yellow. Residues involved in the formation of the binding pocket are represented in stick and coloured according to contributing monomer. Residues which are involved solely in forming the dimer–dimer interface are represented as lines and coloured orange irrespective of contributing monomer.

Figure 4. Channels into the asparagines-binding pocket The channels leading to the ligand-binding pocket of *E.coli* AsnC are illustrated via a section through the molecular surface of the protein with underlying secondary structure elements shown in cartoon representation. The channels into the binding sites from the central cavity are indicated by white dotted lines. The view is looking down the 4-fold axis of the octamer (as in Figure 2e). The asparagine ligands are coloured as carbon atoms in yellow, oxygen atoms in red and nitrogen atoms in blue with the protein surface coloured grey.

binding are heavily dependent on the spatial positioning of L139 to V141. Residues T136 and T138 are highly conserved and their roles in effector binding suggest they constitute an important marker of effector regulation. It is likely that the β5 to αE dimer–dimer contacts would be perturbed by mutations to these residues and would have a significant impact on effector binding by the octamer. The reason for the effect of an I116 mutation (equivalent to M124 in Lrp) is less clear as it is situated 12 Å away from the effector-binding site. However, its location at the dimer–dimer interface and its proximity to sheet β5 suggest it might exert a direct effect upon octamer stability or an indirect longer-range effect on effector binding.

Using a structure-based sequence alignment, mapping of equivalent Lrp residues onto the AsnC structure results in one end of the binding pocket becoming predominantly hydrophobic. The substitution of residues Y82, Y100 and Q128 in AsnC by a phenylalanine and two leucines, respectively, in Lrp, changes the size and nature of the pocket such that it would be ideally suited to binding the aliphatic side chain of leucine as opposed to the carboxyamide group of asparagine (Figure 5). The remainder of the pocket is largely unchanged with the AsnC TET motif (residues 136–138) mimicked by TRT (residues 144–146) in Lrp. Although E137 in AsnC becomes R145 in Lrp, there is a compensatory change in an interacting residue of an H for a D at a position equivalent to residue 104 of AsnC (Figure 1). Furthermore, residue G103 in the turn between β3 and β4 (equivalent to G111 in Lrp) that is critical to the interaction with the generic amino group of the effector is totally conserved across all known sequences (29) and is the only glycine in the structure to possess a positive phi angle (∼90°). Lrp does respond to exogenous alanine and this could clearly be incorporated into the binding pocket. However Lrp does not show a response to Ile or Val and our model would indicate that binding would be disfavoured through possible clashes with the sidechain hydroxyl of T136 and also the backbone carbonyl of T101.

It is also possible to map the pocket residues for the other members of the Lrp/AsnC family for which effector molecules have been identified: FL11 (*Pyrococcus* OT3) binds glutamine (17); PutR (*Agrobacterium tumefaciens*) binds proline (51); BkdR (*Pseudomonas putida*) preferentially binds valine (52); LysM (*S.solfataricus*) exclusively binds lysine (13). In each of these cases the residue changes result in a binding pocket whose steric characteristics and charge distribution fit the proposed ligand. A study of the equivalent region of LrpA finds the binding pocket obstructed by residues M101 and P133 (equivalent to S106 and T138 of AsnC) and is consistent with its reported effector-independent activity. In LrpC, a restructuring of the pocket is achieved via changes to 8 of the 11 residues local to the binding site. The most notable of these are a change in the side chain rotamer of Y78 (equivalent to Y82 in AsnC) and the substitution of Y100 for R96 and Q128 for S123. The resultant binding pocket is smaller in size than that of AsnC, mainly due to the presence of R96, although it still possesses a channel into the central cavity of the octamer. The majority of contacts required to bind the generic carboxyl and amino groups of an amino acid are in place. However, the small size of the pocket precludes the binding of any amino acid much larger than alanine in the absence of any restructuring upon effector binding. The addition of leucine to LrpC does not prevent it binding to its own promoter (25,26) and biochemical studies indicate that LrpC binding to DNA is not influenced by alanine (S. Ayora, unpublished data) and in all likelihood it has no effector. The structure of LrpC contains four water molecules within each of the pockets at the dimer–dimer interface. Changes in the local size and shape of the binding pocket can be modelled based upon our current structural data. However, ligand binding is likely to have larger structural consequences e.g. tertiary and/or quaternary conformational changes, which influence function but cannot be reliably modelled with the available structures.

DNA binding region

Mapping of residues identified in *E.coli* Lrp as being involved in DNA binding [residues D13, L34, L40, S41, P44, L46, R48, Y61, L65 and L70, (50)] onto the structures of AsnC and LrpC reveals two main clusters. Most of the first group of DNA-binding mutants are located in recognition helix αC (Figure 1) and seem likely to be involved in direct contact with the DNA. The possible exceptions are R48 (equivalent to AsnC R42 and LrpC R39), which forms an ion pair network with two totally conserved aspartates in helix αA and L34 (equivalent to AsnC L28 and LrpC L25). These residues map to the inner surface of helix αB and contribute to the core of the domain. The second group of DNA-binding mutants are from strand β1 and are involved in the positioning of helix αA from a second monomer in a dimer pair. Thus these mutations may exert their effect by perturbation of the relative geometry of the helices making up the HtH motif.

Thermostability

The structural elucidation of AsnC and LrpC allows us to investigate potential factors effecting protein stability by comparing them to the structures of LrpA and FL11 from the
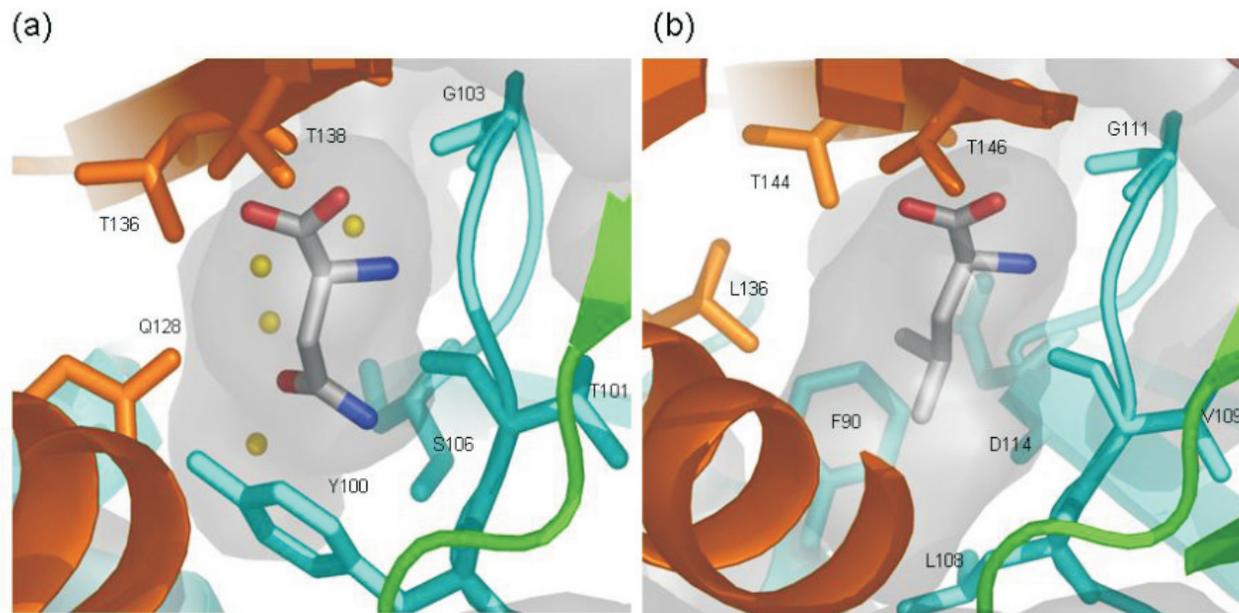
Figure 5. Ligand-binding sites for different amino acid effectors. Schematic representation of the AsnC ligand-binding site in comparison to a model of the Lrp ligand-binding site. (**a**) Stick representation of asparagine bound at the interface between two dimers. Residues contributed by a monomer in the first dimer are shown in brown, with those from monomers in a second dimer pair coloured blue and green. Water molecules are indicated by yellow spheres. (**b**) A model binding site for leucine in *E.coli* Lrp. Five changes were made to the AsnC structure (Y82 to F90; Y100 to L108; T101 to V109; S106 to D114 and Q128 to L136) based on the sequence alignment (Figure 1). The inner surface of the pockets are rendered grey in (a) and (b). In both diagrams the asparagine and leucine are coloured with oxygen atoms in red, nitrogen atoms in blue and carbon in grey.

hyperthermophilic *Pyrococcus* sp. We have analysed the relevant ion pair networks, hydrophobic packing and solvent exposed surface area widely accepted to be associated with increased thermostability in proteins (53–55). With the exception of a triple network found in the N-terminal region of AsnC and LrpC that may contribute to stabilizing the relative orientation of helices αA and αC in the DNA reading heads, only FL11 contained any other ion pair networks with more than two partners (FL11 contained two ion pair networks of four residues in the dimer). LrpA contains the fewest ion pairs and hence this cannot be the reason for the observed thermostability of a protein that has a melting temperature (T_m) of 111.5°C as determined by differential scanning calorimetry (A. B. Brinkman and J. van der Oost, unpublished data). An analysis of the solvent exposed surface areas of the four proteins using AREAIMOL indicates that all four have comparable amounts of surface exposure (ranging from 58 to 66%). Furthermore, a comparison of the number, type and location of hydrophobic amino acids in the mesophilic AsnC/LrpC and hyperthermophilic LrpA/FL11 structures indicates no clear difference that might be used to explain their thermostability. Thus our simple analysis does not reveal a clear explanation based on currently accepted indicators for the basis of the observed thermostability of LrpA and FL11 when compared with their bacterial counterparts.

Implications for DNA binding

It is clear that the N-terminal HtH motif and in particular helix αC is responsible for the binding of DNA in both AsnC and LrpC. The use of such a helical motif to bind DNA is well documented, especially through the work on *E.coli* catabolite

activator protein (CAP) (53,56). The distance between a pair of αC helices in a dimer of either protein might be expected to be 34 Å, which corresponds to the helical periodicity of standard B-form DNA. However, a model of straight B-form DNA bound to a dimer of either AsnC or LrpC suggests that the interactions are not optimal. This is perhaps not surprising given reports that members of the family, such as Lrp, LrpA, LrpC and PutR bind to and induce bending in DNA (6,7,11,25,26). For example, *E.coli* Lrp was found to induce a bend of 52° upon binding to a single site, progressing to a bend of at least 135° upon binding to two adjacent sites (11). The co-crystal structure of the CAP dimer bound to DNA revealed the DNA curvature to be 90°, indicating that binding of two carefully spaced copies of this motif can induce large conformational changes in DNA structure (53). The DNA binding helices in dimers of both AsnC and LrpC (helix αC) adopt a different relative orientation to those of CAP and consequently the distance between the regions of helix αC that contact the DNA are slightly closer. CAP is known to recognize DNA-binding sites separated by ~10 bp or one turn of regular B-form DNA duplex (53). A study of promoter regions regulated by the Lrp/AsnC family revealed several inverted repeat sequences of 5 bp with an intervening 3 bp, i.e. a 13 bp site (54). Thus the centre to centre separation of the DNA-binding heads is only 9 bp for AsnC/LrpC. Analysis of the central 3 bp in the DNA-binding site of AsnC shows a marked preference for TA base pairs, consistent with DNA bending (55). There is also a repeating pattern to the distribution of these 13 bp sites in many of the upstream regions of genes regulated by Lrp/AsnC family members such that either a shorter 7 or 8 bp spacing or a longer 18 bp spacing between sites is generally observed (55).

Modelling of DNA interactions

Modelling of the AsnC octamer bound to a curved piece of DNA (based on PDB code 1CGP) indicates that the pairs of DNA-binding N-terminal domains of individual dimers must be reoriented relative to the effector-binding C-terminal domains to enable binding of an octamer simultaneously to multiple 13 bp sites with 18 bp spacings (Figure 6a). This seems plausible because there are relatively few contacts between the N- and C-terminal domains. Alternatively binding of multiple DNA sites by an array of adjacent octamers might allow the introduction of varying degrees of curvature or even the wrapping of DNA in a solenoid form (Figure 6b). Both AsnC and LrpC have similar diameters of ~118 Å, whereas in contrast LrpA has a diameter of ~107 Å, which makes it difficult to envisage how LrpA could wrap the DNA around itself to accommodate simultaneous binding of a single octamer to more than one site. The majority of the increase in diameter of AsnC and LrpC arises from the packing of the C-terminal residues into the space between elements of strands β3 and β4 and the helical turn around residues P61 to L64 (numbering as for AsnC). This moves helix αA (and consequently the entire N-terminal domain) 3.5 Å further away from the C-terminal core. There is notable sequence conservation in the C-termini of enterobacterial Lrp proteins and this may support a critical role for the C-termini in transmitting the effector-bound status of the protein. There is little difference in the relative orientation of the three helices making up the N-terminal domain, with >94% of Cα atoms superimposing with a root mean square deviation (r.m.s.d.) of 1.4 Å or better over the four known structures. Whilst the exact nature of the interaction of these proteins with their target DNA awaits elucidation, it does appear that the conformation of the C-terminal residues may be important in orienting the position of the HtH domain and thus plays a role in switching preference between differently spaced DNA-binding sites. The importance of such a small shift in structure is supported by work on an *E.coli* Lrp variant lacking the final 11 C-terminal residues which does not bind to *ilvIH* operator DNA, despite having a CD trace comparable to that of the wild-type protein (49).

The effector switch

Like Lrp, AsnC binds DNA and regulates gene expression in the absence or presence of effector. It appears likely that the presence of the effector selects for one possible conformer of the protein rendering it suitable for acting at a subset of target DNA sites but unsuitable for others, presumably by altering the relative orientation and spacing of the N- and C-terminal domains and associated bound DNA. This notion is supported by previous biochemical studies on *E.coli* Lrp and *P.putida* BkdR, which noted a decrease in the intrinsic fluorescence upon their binding of their amino acid effectors to high affinity sites, suggesting a conformational change in both proteins (49,52). Furthermore, slight inter-domain rearrangements have been observed upon effector binding in the structurally analogous ACT domain containing proteins, which are also subject to allosteric regulation by small molecules, typically amino acids [reviewed in (29)]. The exact mechanism by which such a switch could be made is difficult to ascertain in the absence of an effector-free structure of AsnC and awaits elucidation. In addition, in certain cases the presence of the effector may favour an increase in the level of an octameric species of the protein *in vivo* such that a shift in its oligomeric state equilibrium would encourage
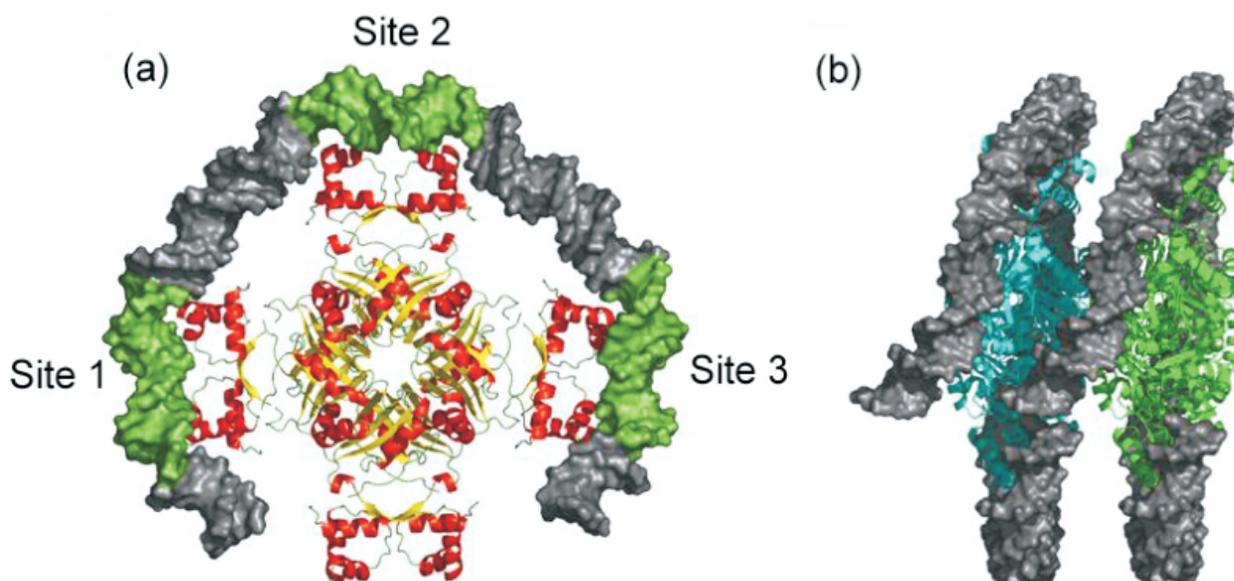
Figure 6. Models of AsnC and LrpC binding to DNA. (**a**) Cartoon representation of AsnC binding to DNA with a curved piece of B-form DNA based on the 22 bp fragment from CRP (53) shown as a surface representation. Dimer-binding sites (labelled 1–3) of 13 bp in length are coloured green and separated by 18 bp of non-conserved DNA (coloured grey) to mimic a promoter region. Binding of DNA reading heads to sites 1 and 3 required a hinge movement around residue 60 to allow a small degree of flexibility between the N- and C-terminal domains of the protein. (**b**) Model of how LrpC could wrap DNA in a nucleosome-like structure. Cooperative binding of LrpC to the DNA forms a right-handed super-helix, which constrains the positive supercoils. Two octamers of the protein are shown in cyan and green. Modelled DNA is shown as a grey surface and based on existing crystal structures of wrapped DNA (PDB code 1AOI) (57) and electron microscopy studies of LrpC (27).

cooperative DNA binding to sites optimized for interaction with an octamer.

DNA binding of the AsnC/Lrp family

Thus it appears that DNA binding in the Lrp/AsnC family is governed by a complex interplay between the oligomeric states populated by the protein and both the number and nucleotide base sequence of mirrored repeats in the upstream regions of the genes concerned. Fine tuning of promoter selection is likely conveyed through subtle changes in structure due to effector binding and/or the packing of the C-terminal residues close to the HtH motif. LrpC reportedly binds DNA in a non-sequence specific manner, recognizing intrinsically curved regions containing phased A-tracts (27). Its stable octameric structure (dimers of the protein were not observed under our solution conditions) indicates that DNA could indeed be wrapped around it to yield a nucleosome-like structure or to influence promoter geometry (Figure 6b), as observed in electron microscopy experiments (27). An extension of the model can be made in which two octamers form a hexadecamer and bind multiple DNA sites, as observed for *E.coli* Lrp (51). This requires a small movement of ∼5 Å and rotation of ∼5° for each of the N-terminal DNA-binding heads relative to the experimentally observed octamer structure such that a bound DNA solenoid structure is formed. The spacing between binding sites is not altered in our model upon the passage of the DNA from one octamer to the next in the hexadecamer.

REFERENCES

1. Charlier,D., Roovers,M., Thia-Toong,T.L., Durbecq,V. and Glansdorff,N. (1997) Cloning and identification of the *Sulfolobus solfataricus lrp* gene encoding an archaeal homologue of the eubacterial leucine responsive global transcriptional regulator Lrp. *Gene*, **201**, 63–68.
2. Napoli,A., Van der Oost,J., Sensen,C.W., Charlebois,R.L., Rossi,M. and Ciaramella,M. (1999) An Lrp-like protein of the hyperthermophilic archaeon *Sulfolobus solfataricus* which binds to its own promoter. *J. Bacteriol.*, **181**, 1474–1480.
3. Brinkman,A.B., Ettema,T.J.G., de Vos,W.M. and van der Oost,J. (2003) The Lrp family of transcriptional regulators. *Mol. Microbiol.*, **48**, 287–294.
4. Willins,D.A., Ryan,C.W., Platko,J.V. and Calvo,J.M. (1991) Characterization of Lrp, an *Escherichia coli* regulatory protein that mediates a global response to leucine. *J. Biol. Chem.*, **266**, 10768–10774.
5. Madhusudhan,K.T., Huang,N. and Sokatch,J.R. (1995) Characterization of BkdR-DNA binding in the expression of the *bkd* operon of *Pseudomonas putida*. *J. Bacteriol.*, **177**, 636–641.
6. Brinkman,A.B., Dahlke,I., Tuininga,J.E., Lammers,T., Dumay,V., de Heus,E., Lebbink,J.H.G., Thomm,M., de Vos,W.M. and van der Oost,J. (2000) An Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus* is negatively autoregulated. *J. Biol. Chem.*, **275**, 38160–38169.
7. Jafri,S., Evoy,S., Cho,K.Y., Craighead,H.G. and Winans,S.C. (1999) An Lrp-type transcriptional regulator from *Agrobacterium tumefaciens* condenses more than 100 nucleotides of DNA into globular nucleoprotein complexes. *J. Mol. Biol.*, **288**, 811–824.
8. Chen,S., Rosner,M.H. and Calvo,J.M. (2001) Leucine-regulated self-association of leucine-responsive regulatory protein (Lrp) from *Escherichia coli*. *J. Mol. Biol.*, **312**, 625–635.
9. D'Ari,R., Lin,R.T. and Newman,E.B. (1993) The leucine-responsive regulatory protein—more than a regulator? *Trends Biochem. Sci.*, **18**, 260–263.
10. Azam,T.A., Iwata,A., Nishimura,A., Ueda,S. and Ishihama,A. (1999) Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J. Bacteriol.*, **181**, 6361–6370.
11. Wang,Q. and Calvo,J.M. (1993) Lrp, a major regulatory protein in *Escherichia coli*, bends DNA and can organize the assembly of a higher-order nucleoprotein structure. *EMBO J.*, **12**, 2495–2501.
12. Peeters,E., Thia-Toong,T.L., Gigot,D., Maes,D. and Charlier,D. (2004) Ss-LrpB, a novel Lrp-like regulator of *Sulfolobus solfataricus* P2, binds cooperatively to three conserved targets in its own control region. *Mol. Microbiol.*, **54**, 321–336.
13. Brinkman,A.B., Bell,S.D., Lebbink,R.J., de Vos,W.M. and van der Oost,J. (2002) The *Sulfolobus solfataricus* Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability. *J. Biol. Chem.*, **277**, 29537–29549.
14. Ouhammouch,M., Dewhurst,R.E., Hausner,W., Thomm,M. and Geiduschek,E.P. (2003) Activation of archaeal transcription by recruitment of the TATA-binding protein. *Proc. Natl Acad. Sci. USA*, **100**, 5097–5102.
15. Ouhammouch,M., Langham,G.E., Hausner,W., Simpson,A.J., El-Sayed,N.M.A. and Geiduschek,E.P. (2005) Promoter architecture and response to a positive regulator of archaeal transcription. *Mol. Microbiol.*, **56**, 625–637.
16. Bell,S.D. and Jackson,S.P. (2001) Mechanism and regulation of transcription in archaea. *Curr. Opin. Microbiol.*, **4**, 208–213.
17. Koike,H., Ishijima,S.A., Clowney,L. and Suzuki,M. (2004) The archaeal feast/famine regulatory protein: potential roles of its assembly forms for regulating transcription. *Proc. Natl Acad. Sci. USA*, **101**, 2840–2845.
18. Tani,T.H., Khodursky,A., Blumenthal,R.M., Brown,P.O. and Matthews,R.G. (2002) Adaptation to famine: a family of stationary-phase genes revealed by microarray analysis. *Proc. Natl Acad. Sci. USA*, **99**, 13471–13476.
19. Calvo,J.M. and Matthews,R.G. (1994) The leucine responsive regulatory protein, a global regulator of metabolism in *Escherichia coli*. *Microbiol. Rev.*, **58**, 466–490.
20. Newman,E.B. and Lin,R.T. (1995) Leucine-responsive regulatory protein – a global regulator of gene expression in *Escherichia coli*. *Ann. Rev. Microbiol.*, **49**, 747–775.
21. Willins,D.A., Ryan,C.W., Platko,J.V. and Calvo,J.M. (1991) Characterization of Lrp, an *Escherichia coli* regulatory protein that mediates a global response to leucine. *J. Biol. Chem.*, **266**, 10768–10774.
22. Kolling,R. and Lother,H. (1985) AsnC—an autogenously regulated activator of asparagine synthetase A transcription in *Escherichia coli*. *J. Bacteriol.*, **164**, 310–315.
23. de Wind,N., de Jong,M., Meijer,M. and Stuitje,A.R. (1985) Site directed mutagenesis of the *Escherichia coli* chromosome near *oriC*: identification and characterization of *asnC*, a regulatory element in *E.coli* asparagine metabolism. *Nucleic Acids Res.*, **13**, 8797–8811.
24. Beloin,C., Ayora,S., Exley,R., Hirschbein,L., Ogasawara,N., Kasahara,Y., Alonso,J.C. and LeHegarat,F. (1997) Characterization of an *lrp*-like (*lrpC*) gene from *Bacillus subtilis*. *Mol. Gen. Genet.*, **256**, 63–71.
25. Beloin,C., Exley,R., Mahe,A.L., Zouine,M., Cubasch,S. and Le Hegarat,F. (2000) Characterization of LrpC DNA-binding properties and regulation of *Bacillus subtilis lrpC* gene expression. *J. Bacteriol.*, **182**, 4414–4424.

26. Tapias,A., Lopez,G. and Ayora,S. (2000) *Bacillus subtilis* LrpC is a sequence-independent DNA-binding and DNA-bending protein which bridges DNA. *Nucleic Acids Res.*, **28**, 552–559.

27. Beloin,C., Jeusset,J., Revet,B., Mirambeau,G., Le Hegarat,F. and Le Cam,E. (2003) Contribution of DNA conformation and topology in right-handed DNA wrapping by the *Bacillus subtilis* LrpC protein. *J. Biol. Chem.*, **278**, 5333–5342.

28. Leonard,P.M., Smits,S.H.J., Sedelnikova,S.E., Brinkman,A.B., de Vos,W.M., van der Oost,J., Rice,D.W. and Rafferty,J.B. (2001) Crystal structure of the Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus. EMBO J.*, **20**, 990–997.

29. Ettema,T.J.G., Brinkman,A.B., Tani,T.H., Rafferty,J.B. and van der Oost,J. (2002) A novel ligand-binding domain involved in regulation of amino acid metabolism in prokaryotes. *J. Biol. Chem.*, **277**, 37464–37468.

30. Platko,J.V. and Calvo,J.M. (1993) Mutations affecting the ability of *Escherichia coli* Lrp to bind DNA, activate transcription, or respond to leucine. *J. Bacteriol.*, **175**, 1110–1117.

31. Schneider,T.R. and Sheldrick,G.M. (2002) Substructure solution with SHELXD. *Acta Crystallogr. D*, **58**, 1772–1779.

32. Hendrickson,W.A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*, **254**, 51–58.

33. Leslie,A.G.W. (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 and ESF-EAMCB Newsletter on Protein Crystallography*, **26**.

34. Evans,P.R. Scaling of MAD data. In *Proceedings of the CCP4 Study Weekend. Recent Advances in Phasing*, 97–102.

35. Roussel,A., Fontecilla-Camps,J.C. and Cambillau,C. (1990) TURBO-FRODO: a new program for protein crystallography and modelling. *XV IUCr Congress Abstracts*, 66–67.

36. Bricogne,G., Vonrhein,C., Flensburg,C., Schiltz,M. and Paciorek,W. (2003) Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D*, **59**, 2023–2030.

37. Abrahams,J.P. and Leslie,A.G.W. (1996) Methods used in the structure determination of bovine mitochondrial F-1 ATPase. *Acta Crystallogr. D*, **52**, 30–42.

38. Collaborative Computational Project, Number 4. (1994), The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.

39. Vagin,A. and Teplyakov,A. (2000) An approach to multi-copy search in molecular replacement. *Acta Crystallogr. D*, **56**, 1622–1624.

40. Cowtan,K.D. (1994) An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EAMCB Newsletter on Protein Crystallography*, **31**, 34–38.

41. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D*, **60**, 2126–2132.

42. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D*, **53**, 240–255.

43. Laskowski,R.A., Macarthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK—a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

44. Perrakis,A., Morris,R. and Lamzin,V.S. (1999) Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.*, **6**, 458–463.

45. Nicholls,A., Bharadwaj,R. and Honig,B. (1993) GRASP—graphical representation and analysis of surface-properties. *Biophys. J.*, **64**, A166–A166.

46. Aravind,L., Anantharaman,V., Balaji,S., Babu,M.M. and Iyer,L.M. (2005) The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *FEMS Microbiol. Rev.*, **29**, 231–262.

47. Lee,B. and Richards,F.M. (1971) The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.*, **55**, 379–400.

48. Jones,S. and Thornton,J.M. (1995) Protein–protein interactions—a review of protein dimer structures. *Prog. Biophys. Mol. Biol.*, **63**, 31–65.

49. Chen,S. and Calvo,J.M. (2002) Leucine-induced dissociation of *Escherichia coli* Lrp hexadecamers to octamers. *J. Mol. Biol.*, **318**, 1031–1042.

50. Platko,J.V. and Calvo,J.M. (1993) Mutations affecting the ability of *Escherichia coli* Lrp to bind DNA, activate transcription, or respond to leucine. *J. Bacteriol.*, **175**, 1110–1117.

51. Cho,K.Y. and Winans,S.C. (1996) The *putA* gene of *Agrobacterium tumefaciens* is transcriptionally activated in response to proline by an Lrp-like protein and is not autoregulated. *Mol. Microbiol.*, **22**, 1025–1033.

52. Madhusudhan,K.T., Huang,N., Braswell,E.H. and Sokatch,J.R. (1997) Binding of L-branched-chain amino acids causes a conformational change in BkdR. *J. Bacteriol.*, **179**, 276–279.

53. Schultz,S.C., Shields,G.C. and Steitz,T.A. (1991) Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science*, **253**, 1001–1007.

54. Suzuki,M. (2003) The DNA-binding specificity of eubacterial and archaeal FFRPs. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.*, **79**, 213–222.

55. Koo,H.S., Wu,H.M. and Crothers,D.M. (1986) DNA bending at adenine:thymine tracts. *Nature*, **320**, 501–506.

56. McKay,D.B. and Steitz,T.A. (1981) Structure of catabolite gene activator protein at 2.9Å resolution suggests binding to left-handed B-DNA. *Nature*, **290**, 744–749.

57. Luger,K., Mader,A.W., Richmond,R.K., Sargent,D.F. and Richmond,T.J. (1997) Crystal structure of the nucleosome core particle at 2.8Å resolution. *Nature*, **389**, 251–260.