Crystal structure of the sensory domain of *Escherichia coli* CadC, a member of the ToxR-like protein family

Andreas Eichinger,¹ Ina Haneburger,² Christiane Koller,² Kirsten Jung,² and Arne Skerra^{1*}

¹Munich Center for Integrated Protein Science and Lehrstuhl für Biologische Chemie, Technische Universität München, Freising-Weihenstephan, Germany

²Munich Center for Integrated Protein Science and Department of Biology I, Microbiology, Ludwig-Maximilians-Universität, München, Germany

Received 23 September 2010; Revised 27 December 2010; Accepted 9 January 2011

DOI: 10.1002/pro.594

Published online 24 January 2011 proteinscience.org

Abstract: The membrane-integral transcriptional activator CadC comprises sensory and transcriptional regulatory functions within one polypeptide chain. Its C-terminal periplasmic domain, CadC_{pd}, is responsible for sensing of environmental pH as well as for binding of the feedback inhibitor cadaverine. Here we describe the crystal structure of CadC_{pd} (residues 188–512) solved at a resolution of 1.8 Å via multiple wavelength anomalous dispersion (MAD) using a ReCl₆^{2−} derivative. CadC_{pd} reveals a novel fold comprising two subdomains: an N-terminal subdomain dominated by a β-sheet in contact with three α-helices and a C-terminal subdomain formed by an eleven-membered α-helical bundle, which is oriented almost perpendicular to the helices in the first subdomain. Further to the native protein, crystal structures were also solved for its variants D471N and D471E, which show functionally different behavior in pH sensing. Interestingly, in the heavy metal derivative of CadC_{pd} used for MAD phasing a ReCl₆^{2−} ion was found in a cavity located between the two subdomains. Amino acid side chains that coordinate this complex ion are conserved in CadC homologues from various bacterial species, suggesting a function of the cavity in the binding of cadaverine, which was supported by docking studies. Notably, CadC_{pd} forms a homodimer in solution, which can be explained by an extended, albeit rather polar interface between two symmetry-related monomers in the crystal structure. The occurrence of several acidic residues in this region suggests protonation-dependent changes in the mode of dimerization, which could eventually trigger transcriptional activation by CadC in the bacterial cytoplasm.

Keywords: cadaverine; pH-sensing; periplasmic domain; ToxR-like protein; X-ray structure

Introduction

Bacteria adapt to changes of external pH by various physiological processes and alterations in gene

expression. Thus, several genes are induced in the neutrophilic bacterium *Escherichia coli* upon exposure to acidic conditions to maintain the cytoplasmic pH in the physiological range between 7.6 and 7.8.¹ In particular, the degradative amino acid decarboxylase systems *adi*, *gad*, and *cad* are among those genes.² Under conditions of an acidic environment and in the presence of external lysine the membrane-integral transcriptional activator CadC is triggered and induces transcription of the *cadBA* operon, which encodes the cadaverine/lysine-antiporter CadB and the lysine decarboxylase CadA. CadA intracellularly decarboxylates lysine to cadaverine

Christiane Koller's current address is Institute of Physiological Chemistry, Faculty of Veterinary Medicine, Ludwig-Maximilians-Universität, Munich, Germany.

Grant sponsor: the Deutsche Forschungsgemeinschaft; Grant numbers: Exc114/1 and JU270/5-3.

*Correspondence to: Prof. Dr. Arne Skerra, Lehrstuhl für Biologische Chemie, Technische Universität München, Emil-Erlenmeyer-Forum 5, 85350 Freising-Weihenstephan, Germany. E-mail: skerra@tum.de

under consumption of one proton. The resulting cadaverine is exported by CadB, which in turn takes up lysine from the external medium.³

CadC is an integral membrane protein of 512 amino acids comprising an N-terminal cytoplasmic DNA-binding domain (residues 1–159), a transmembrane helix (residues 160–187), and a C-terminal periplasmic domain (residues 188–512).⁴ CadC belongs to the family of ToxR-like proteins, which includes CadC of *Salmonella enterica* serovar Typhimurium and *Vibrio cholerae* (regulating the *cadBA* operon),⁵ ToxR from *V. cholerae* (regulating cholera toxin, pilus, and outer-membrane protein expression),⁶ TcpP from *V. cholerae* (ToxR-coregulator of virulence gene expression),⁷ PsaE from *Yersinia pestis* (regulating fimbriae genes),⁸ and WmpR from *Pseudoalteromonas tunicata* (regulating iron uptake).⁹ Despite low sequence homology (outside the CadC orthologs) proteins of this family are characterized by a common three-domain topology and all combine sensory and transcriptional regulatory functions within one polypeptide chain. Therefore, these proteins represent the simplest known transmembrane signaling system that transduces information across the lipid bilayer without involving chemical modification.

The DNA-binding function of CadC is mediated by its cytoplasmic domain that contains a helix-turn-helix motif similar to the RO_{II} subgroup of DNA-binding domains found in response regulators such as PhoP from *Bacillus subtilis*,¹⁰ VirG from *Agrobacterium tumefaciens*¹¹ or OmpR from *Escherichia coli*.¹² The DNA-binding sites were identified and the interaction between purified CadC and the promoter region was demonstrated *in vitro*.¹³ Expression of the *cadBA* operon is activated at low external pH and concomitantly available lysine. Recently, it was shown that CadC is not a direct sensor of the external lysine concentration. Instead, lysine is co-sensed via interplay with the lysine-specific permease LysP.^{14,15} Still, the pH-sensory function is assigned to the periplasmic domain of CadC.^{16,17} Several CadC variants with single amino acid replacements in the periplasmic domain affecting the pH-dependent *cadBA* expression were previously identified.^{16,17} For example, replacement of Asp471 against Asn resulted in a pH-independent activation of expression. In contrast, CadC with Glu at the same position was no longer able to respond to acidification of the environment. Thus, single amino acid replacements can fix CadC either in an ON or an OFF state, respectively.

A combination of biochemical experiments and computational simulations revealed that the transcriptional stimulating activity of CadC is further regulated via feedback inhibition by the product of the CadA decarboxylation reaction, namely cadaverine.¹⁸ Consistently, cadaverine binds to the periplasmic domain of CadC with moderate affinity ($K_D =$

96 μ M).¹⁴ Nevertheless, it is hardly understood how CadC adopts its active state after receiving the pH signal and, also, how cadaverine inactivates CadC. To gain insight into the structural mechanisms, we have carried out a crystallographic analysis of the periplasmic domain CadC_{pd} and its two variants D471E and D471N.

Results

Three-dimensional structure of the periplasmic domain of CadC

The X-ray structure of CadC_{pd} (residues 188–512) was solved via multiple wavelength anomalous dispersion (MAD) at a resolution of 2.3 Å by using a crystal soaked with K₂ReCl₆. Further crystal structures were solved for the native apo-protein at a resolution of 1.8 Å as well as for the CadC_{pd} variants D471E and D471N at resolutions of 1.9 and 2.2 Å, respectively (Table I). CadC_{pd} appears as an irregularly shaped protein with a maximal diameter of about 60 Å and a diameter of approximately 30 Å at its narrowest site. The entire periplasmic protein domain is essentially composed of two subdomains: the first subdomain comprises a mixed, parallel/antiparallel β -sheet packed on one side against two α -helices as part of a three-helix bundle, whereas the second subdomain is a pure α -helical bundle encompassing 11 helices (Fig. 1). Taken together, α -helices and β -strands account for 70% of the 314 ordered residues in the periplasmic domain of the wild type CadC apo-protein.

Following to an N-terminal loop at positions 191–194, the polypeptide chain first enters a five-stranded β -sheet at strand B. The chain then adopts a flexible loop not fully defined in the electron density, runs back through an α -helix (helix 1) and enters the β -sheet again as strand A, which is arranged parallel to strand B. After a long loop the chain proceeds back to the β -sheet as strand C, again parallel to B. Two more strands, D and E, connected by shorter loops, follow and are aligned antiparallel to strand C and to each other. The chain then enters an element of two antiparallel α -helices (helices 2 and 3) which, together with helix 1, form the three-helix bundle in the N-terminal subdomain. Notably, all secondary structure elements within this subdomain are oriented in the same direction. For residue Cys208, which is part of the flexible loop connecting strand A and helix 1, no electron density was observed with any dataset, indicating a presumably free (reduced) state in these crystal structures. Because of its proximity to the only other thiol side chain of the well resolved residue Cys272, however, a disulfide bond could be formed between these two amino acids and, for the purpose of illustration, has been modelled here [cf. Fig. 1(A)]. Indeed, the presence of such a disulfide bridge was detected for

Table I. Crystallographic and Refinement Statistics

Protein	CadC _{pd} • ReCl ₆ ²⁻			CadC _{pd}		
Dataset	Peak	Inflection	Remote	Wild Type	D471E	D471N
Space group		P6 ₁ 22		P6 ₁ 22	P6 ₁ 22	P3 ₂ 21
Unit cell dimensions, a, b, c [Å], α = β = 90°, γ = 120°	80.56 80.56 199.78	80.88 80.88 200.30	81.00 81.00 200.65	83.90 83.90 199.09	83.63 83.63 199.49	79.73 79.73 126.01
Wavelength [Å]	1.17652	1.17705	1.07813	0.91841	0.91841	0.91841
Resolution range [Å] ^a	69.77–2.30 (2.36–2.30)	70.04–2.60 (2.74–2.60)	70.15–2.90 (3.06–2.90)	72.66–1.80 (1.85–1.80)	72.43–1.90 (1.95–1.90)	69.05–2.20 (2.26–2.20)
I/σ ^a	6.2 (2.1)	7.4 (2.2)	9.0 (2.5)	5.6 (2.7)	5.2 (2.6)	20.2 (4.6)
R _{merge} ^{a,b} [%]	9.5 (35.5)	8.0 (34.1)	6.7 (30.6)	8.5 (27.6)	8.6 (27.6)	5.2 (36.7)
Unique reflections	16,693	12,472	9,169	39,276	33,530	24,043
Multiplicity ^a	14.1 (14.5)	13.9 (14.4)	13.8 (14.3)	14.3 (15.1)	13.5 (13.7)	5.6 (5.1)
Completeness ^a	98.5 (97.9)	99.2 (98.8)	99.4 (99.2)	99.9 (100.0)	99.9 (100.0)	99.6 (99.9)
Anomalous multiplicity ^a	7.7 (7.7)	7.7 (7.7)	7.7 (7.7)	–	–	–
Anomalous completeness ^a	99.5 (98.9)	99.8 (99.5)	99.9 (99.7)	–	–	–
Phasing and refinement: Figure of merit centric, acentric reflections		0.697, 0.419		–	–	–
R _{cryst} , R _{free} ^c [%]		23.0, 27.1		19.1, 21.8	20.2, 23.1	22.8, 27.9
Protein atoms		2,268		2,497	2,507	2,362
Ligand atoms		112		–	–	–
Water molecules		34		237	234	114
Average B factor [Å ²]		24.34		17.04	22.38	33.60
Geometry: R.m.s.d. bond lengths, angles [Å, °]					0.014, 1.300	0.018, 1.513
Ramachandran analysis: Core, allowed, generously allowed, disallowed [%]		91.3, 8.3, 0.0, 0.4		93.0, 6.7, 0.0, 0.4	93.4, 5.9, 0.3, 0.3	93.2, 6.1, 0.4, 0.4

^a Values in parentheses are for the highest resolution shell.

^b R_{free} is the R_{cryst} with 5% of the reflections that were randomly selected and excluded from refinement: $R_{\text{cryst}} = \sum_h ||F_o(h)|| / \sum_h ||F_o(h)||$

^c $R_{\text{merge}} = \sum_h \sum_i |I_i(h) - \langle I(h) \rangle| / \sum_h \sum_i I_i(h)$

CadC_{pd} in biochemical experiments (Tetsch *et al.*, in preparation).

Separated by a loop that forms a sharp turn between residues Leu330 and Leu336, the polypeptide chain enters the C-terminal subdomain of CadC_{pd}. Its altogether 11 α -helices (nos. 4–14) are organized as a

bundle of five antiparallel α -helical pairs, whereby helix 13 comprises just one turn. These pairs are twisted in a clockwise direction, resulting in a spiral bundle such that the C-terminus of helix 4 is situated close to the N-terminal end of helix 12. The very short helix 13 together with helix 14 form the C-terminus of the structure. Notably, the bundle of α -helices in the second subdomain is roughly oriented perpendicular to the three-helix motif of the first subdomain.

Comparison between the CadC_{pd} apo-structure and its rhenate complex

The crystal structures of the apo-form of CadC_{pd} and its rhenate complex are very similar even though a rather large number of altogether 16 bound ReCl₆²⁻ complex ions was identified: the root mean square deviation (r.m.s.d.) of the polypeptide chains was 0.79 Å upon superposition of the 282 C α atoms of common ordered residues in all four models (192–204, 217–245, 251–261, 273–293, 298–394, and 399–509). However, three exposed regions showed significant structural differences [cf. Fig. 1(B)]. One deviation occurs at a stretch formed by residues 294–300, which are part of a loop at the N-terminus of α -helix 2. Second, the stretch 331–339 forms part of a loop at the N-terminus of α -helix 4 and is located close to a bound ReCl₆²⁻ ion at the protein surface. Finally, the segment 388–399 forms a loop connecting α -helices 6 and 7, whose C α positions are shifted up to 6.7 Å. Two ReCl₆²⁻ ions were bound in the vicinity of this site.

Structures of the CadC_{pd} variants D471E and D471N

Residue Asp471 was previously identified in a random mutagenesis study to be involved in pH-sensing.¹⁶ Recent site-directed mutagenesis experiments

Figure 1. Overview of the CadC_{pd} crystal structure. (A) Cartoon representation of the apo-protein solved at 1.8 Å. The protein consists of two subdomains. Secondary structure elements are depicted in yellow (β -strands), blue (α -helices of the N-terminal subdomain), and salmon (α -helices of the C-terminal subdomain). The two short 3_{10} -helices 334–336 and 411–413 are also shown in salmon. The disordered loop 207–214 has been modelled and is depicted in light gray. The disulfide bond connecting residues Cys208 and Cys272 is shown as red sticks and labeled. (B) Superposition of the CadC_{pd} apo-protein (yellow-orange) with the ReCl₆²⁻ complex (raspberry) and the variants D471E (green) and D471N (blue). Disordered loops were modelled with plausible geometry and are depicted in light gray. (C) Electrostatic surface representation of the CadC_{pd} apo-structure in an orientation identical to (B). Negatively charged areas are colored red ($-10 k_B T/e$), positively charged areas are colored blue ($10 k_B T/e$). (D) The molecule in the same representation rotated vertically by around 180° with respect to (C), illustrating a narrow entrance to the central cavity indicated by a dashed ellipse. Residues mentioned in the text are labeled. An [interactive view](#) is available in the electronic version of the article.

furthermore revealed that replacement of Asp471 with Glu (D471E) resulted in a CadC variant preventing *cadBA* expression, also under acidic conditions (OFF state), whereas the variant with an Asn side chain at this position (D471N) led to a pH-insensitive active protein (ON state).¹⁷ The X-ray structures of wild type CadC_{pd} and its variant D471E were superimposable with an r.m.s.d. value of 0.11 Å using 282 Cα atoms, indicating a virtually identical conformation [Fig. 1(B)]. However, the deviations in the variant D471N, which was crystallized under acidic conditions in a different space group (see Table I), were more significant and reflected by a higher r.m.s.d. value of 0.95 Å.

Considerable differences in the structures of wild type CadC_{pd} and its variant D471N occur at residues 296–297 in a loop between β-strand E and α-helix 2 in the N-terminal subdomain. There is also discrepancy in a long stretch comprising residues 320–340, which connects the N- and the C-terminal subdomains. It includes α-helix 3, a loop followed by a 3₁₀-helix, and the N-terminal part of α-helix 4 [cf. Fig. 1(B)]. Another deviation is evident at the amino acid stretch 388–415 comprising the C-terminus of α-helix 6, a loop, α-helix 7, another loop, a 3₁₀-helix, and the connecting loop to α-helix 8. In contrast, α-helices 9–14 do not show significant deviations, in particular α-helix 11, which carries the mutation at position 471. Notably, among all those differences only two loops are involved in crystal packing contacts: stretch 293–298 with contacting residues Val293, Thr295, Asn296, and Asn298 and stretch 390–393 with contacting residues His390 and Pro391. Overall, it seems that wild type CadC_{pd} was crystallized in an inactive state also mimicked by CadC_{pd}(D471E), whereas CadC_{pd}(D471N) should better represent the active state, which may also be reflected by the fact that the D471N mutation led to a different crystallization behavior (and altered space group) of CadC_{pd}.

Crystal contacts of the CadC_{pd} wild type protein and its variants D471E and D471N

The wild type CadC_{pd} (both apo-protein and Re-complex) and the variant D471E isomorphously crystal-

lized in the same space group P6₁22 and, hence, both proteins form the same crystal packing contacts with altogether five neighboring symmetry-equivalent molecules. Interestingly, a detailed analysis of these packing interactions using the PISA server¹⁹ indicated the formation of a functional dimer with one of the neighboring molecules, which was similarly detectable also in the P3₂21 crystal form of the D471N variant (Fig. 2). The putative dimers in all three crystal structures yielded a complexation significance score (CSS) of 1.000. Furthermore, the shape complementarity of the wild type CadC dimer was calculated as $S_C = 0.76$ (using a 1.7 Å probe sphere radius), which is in the range of physiologically functional protein dimers.²⁰ These structural findings were backed by biochemical results of size exclusion chromatography experiments, revealing dimerization of the periplasmic domain also in solution [cf. Fig. 2(D)].

The dimer formation can be described as an extended contact between two more or less planar and surprisingly polar surfaces on both monomers with local C2 symmetry. In addition, each subunit forms a short arm with its N-terminal loop residues 190–196, which nestles in a crevice of the opposite subunit. Loop residues 248–252 form one side of this niche and residues 504–511 from the C-terminal α-helix 14 form the other side [cf. Fig. 2(A)]. The total dimer interface of the wild type apo-protein buries a solvent accessible area of 1,761 Å² per monomer, which corresponds to 11.2% of the total molecular surface. The dimer interface comprises 53 residues from each monomer with an extensive network of 11 hydrogen bonds and 6 salt bridges, including 50 buried water molecules, but exhibits only few (20) hydrophobic side chains. In case of CadC_{pd}(D471E) the interface area is slightly larger with 1,994 Å² (12.6% of the total molecular surface) comprising 58 residues with 11 hydrogen bonds and 2 salt bridges. However, the interface area of CadC_{pd}(D471N) is just marginally smaller, with 1,732 Å² (10.8%), than the one of the wild type protein, comprising 51 residues with 9 hydrogen bonds and 2 salt bridges.

Figure 2. Analysis of CadC_{pd} dimerization. (A) Stereo view of one CadC_{pd} dimer related by a crystallographic symmetry axis. One monomer is colored as in Figure 1(A), the other one is uniformly colored cyan. The N- and C-termini are both directed towards the lipid bilayer indicated at the bottom. (B) Hydrophobic surface representation of the two CadC_{pd} monomers rotated by 90° about a horizontal axis (with the N-termini directed toward the front) and then rotated by ±90° apart to reveal the dimer interface. The interface forming residues are colored according to increasing hydrophobicity from green (hydrophilic) over white (neutral, including the polypeptide backbone) to brown (hydrophobic).⁴⁷ Residues Asp471 and Lys255, which form a salt bridge, are labeled. (C) The same view as in (B) with electrostatic surface representation as in Figure 1(C and D). (D) Analytical size exclusion chromatography of the recombinant CadC_{pd} at pH 7.5. The protein eluted with a retention volume of 14.4 mL, which corresponds to a molecular size of 64.8 kDa (left) as determined from a plot of log MW of calibration proteins against their elution volumes (right): cytochrome c (12.4 kDa), carbonic anhydrase (29 kDa), ovalbumin (43 kDa), bovine serum albumin (66.3 kDa), alcohol dehydrogenase (150 kDa), β-amylase (200 kDa), apoferritin (443 kDa). The apparent size of CadC_{pd} is considerably larger than the calculated mass of the proteolytic product after thrombin cleavage (41.8 kDa), indicating stable dimer formation in solution.

Interestingly, all the different side chains at position 471, that is Asp, Glu, or Asn, participate in a (pairwise symmetric) dimer contact. The wild type

residue Asp471 forms several hydrogen bonds with its carboxylate group, both to two water molecules and to the NZ atom of Lys255 in the neighboring

Figure 2.

molecule. In the isomorphously crystallized variant D471E both carboxylate oxygens of the longer Glu471 side chain form direct hydrogen bonds to the NZ atom of Lys255. In the variant D471N the mutated residue Asn471 forms one direct hydrogen bond to Lys255 NZ and a second one to a water molecule. Apart from residue Asp471, there are up to seven other acidic amino acids involved in the dimer interface: Asp198, Asp248, Glu249, Asp280, Asp283, Asp445, and Glu468. Of these, Glu249 and Asp283 form salt bridges with residues Arg191 and Arg467, respectively. Such a polar and wet, loosely packed, and both-sided negatively charged interface is unusual for a stable protein dimer. In fact, these peculiar biophysical properties suggest a charge-driven sensing mechanism of the CadC transcriptional activator as will be further discussed below.

The putative ligand cavity and the ReCl_6^{2-} complex of CadC_{pd}

The two subdomains of CadC_{pd} enclose a cylinder-shaped cavity in between, which is about 20 Å long and has a diameter of 13 Å. A narrow hole, lateral to the dimer interface described above, gives entrance to this cavity [Fig. 1(D)] from the solvent. Its diameter is about 5 Å and it is lined by residues Arg321, Glu324, Lys328, Asn414, and Asn415. Side chains of altogether 46 residues form the central cavity in CadC_{pd} . Its internal surface is made by 20 aliphatic residues, of which 11 side chains (Leu347, Leu348, Leu355, Leu381, Val385, Val425, Leu428, Leu451, Val454, Leu455, and Ile493) form a hydrophobic pocket within the α -helical subdomain. Ten residues contribute polar side chains, of those Arg321 and Arg480 are positively charged and Asp225, Asp241, Glu378, and Glu447 are negatively charged [Fig. 3(A)]. Notably, in the heavy atom derivative that had been prepared for phase determination the cavity harbored one of the ReCl_6^{2-} complex ions [Fig. 3(C)] while in the CadC_{pd} apo-structure the cavity was filled with a cluster of 20 ordered solvent molecules arranged in typical hydrogen bonding distances.

Considering the in total doubly negative charge of the hexachlororhenate ion it is surprising that this ligand became accommodated in the mostly negatively charged cavity. Yet, the metal-coordinating chloride ions might be polarized by the electron-deficient Re^{IV} central ion to such an extent that positive partial charges appear at the surface of this complex ion. Apart from the electrostatic aspect, the ReCl_6^{2-} sterically fits well into the cavity and forms close contacts with five residues: Asp225, Thr229, His344, Glu447, and Trp450. In fact, their polar groups are in hydrogen bonding distance to the six chloride ions. In addition, there are three water molecules located in the cavity that contribute hydrogen bonds: one water molecule contacts Thr229, Asn232, and Met448, the second one contacts Ser417 and Ile418,

and the third one forms hydrogen bonds to residues Gln421 and Glu447.

Docking studies with cadaverine

In principle, the cavity between the two subdomains of the periplasmic region would be nicely suited to accommodate the positively charged cadaverine ligand of CadC, both with regard to sterical fit and to electrostatic complementarity. Unfortunately, soaking experiments with this ligand followed by crystallographic analysis did not give rise to well defined electron density. To identify side chains that could possibly interact with cadaverine within the presumed binding site, we performed a computer simulation using the program AutoDock²¹ with cadaverine initially provided in a random orientation. All the 10 best scoring results yielded cadaverine in a mostly extended conformation and bound with similar orientation in the cavity [Fig. 3(D)]. Cadaverine was surrounded by residues Glu447, Gln421, and Tyr374 on one side and Asp225 as well as Thr229 on the other. The polar or negatively charged side chains were appropriately positioned to form hydrogen bonds to both positively charged amino groups of the ligand. AutoDock reported calculated dissociation constants for these 10 best CadC_{pd} -cadaverine complexes in the range between 12 and 28 μM , which is in the same order of magnitude as the K_{D} values from $96 \pm 18 \mu\text{M}$ previously measured for the wild type protein via fluorescence titration.¹⁴

Comparison between CadC_{pd} and proteins with related folds

A search for structurally related proteins in the protein data bank (PDB) with DALI²² revealed several hits with Z-scores ranging from 11.8 to 9.5. Among the best 20 candidates no protein with a recognizable structural homology was found. Only parts of two structures (hit nos. 8 and 15) showed some local similarity with the α -helical subdomain of CadC_{pd} (Fig. 4). The peroxisomal targeting signal 1 binding domain of *Trypanosoma brucei* peroxin 5²³ (PDB entry 3CVN; Z-score = 10.6) exhibits a twisted 18 α -helix bundle. A fragment of this structure formed by α -helices 3 to 12 shows some resemblance with CadC_{pd} [Fig. 4(B)]. However, it is less compact and the mutual arrangement of the α -helices is more entwined in this section while one of the α -helices clearly has a different orientation. The transcription factor MalT domain III of *Escherichia coli*²⁴ (PDB entry 1HZ4; Z-score = 9.9) likewise forms a twisted 18 α -helix structure. A fragment comprising α -helices 6–15 again shows similarity with the α -helical subdomain of CadC_{pd} . Interestingly, this protein is also a transcriptional activator. The other 18 hits were predominantly all- α -helical proteins but did not exhibit significant similarity with CadC_{pd} . Another search for structurally related proteins using the

HHPred server²⁵ revealed one additional protein with partial similarity to CadC_{pd}: the tetratricopeptide repeat protein NlpI of *Escherichia coli* (PDB entry 1XNF).²⁶ A fragment of this lipoprotein comprising α -helices 4 to 13 grossly resembles the α -helical architecture of CadC_{pd}, yet the helix arrange-

ment is clearly more open. Consequently, CadC_{pd} exhibits a novel fold, both with the 11- α -helix bundle and the mixed β -sheet/ α -helix assembly in its two subdomains.

Homologous proteins with similar sequence motifs

A BLAST²⁷ search with the amino acid sequence of CadC_{pd} in the NCBI nonredundant protein sequences database²⁸ revealed homologous CadC proteins from 24 different bacterial species with sequence identities ranging from 91 to 21%, namely (with GI numbers of the NCBI database given in parentheses) *Escherichia albertii* (170766700), *Escherichia fergusonii* (218549242), *Klebsiella variicola* (288936734), *Klebsiella pneumoniae* (206578436), *Salmonella enterica* (161502288), *Yersinia ruckeri* (238756479), *Salmonella typhimurium* (16765877), *Serratia proteamaculans* (15737200), *Serratia odorifera* (293394724), *Edwardsiella tarda* (294635142), *Edwardsiella ictaluri* (238918793), *Vibrio cholerae* (147674140), *Vibrio orientalis* (261250472), *Vibrio harveyi* (156972567), *Vibrio shilonii* (149190651), *Vibrio metschnikovii* (260774390), *Vibrio vulnificus* (27365392), *Vibrio mimicus* (262163818), *Vibrio parahaemolyticus* (28899667), *Vibrio fischeri* (59712667), *Vibrio alginolyticus* (91228716), *Aeromonas hydrophila* (117620838), *Aliivibrio salmonicida* LFI1238 (209695912), and *Acidobacteria bacterium* Ellin345 (94968198). Apart from these hits there were only two shorter sequences with similarity to CadC_{pd}, yet corresponding to proteins of other classes: the putative serine/threonine kinase of *Psychroflexus torquis* (91215656) with 24% identity to the CadC sequence 266–381 and the TPR repeat-containing serine/threonine protein kinase of

Figure 3. Analysis of the cavity and the interface between the two subdomains of CadC_{pd}. (A) Electrostatic surface properties (cf. Fig. 1) of the C-terminal α -helical subdomain (residues 333–509; left) and the N-terminal predominantly β -sheet subdomain (residues 190–332; right). Both moieties were rotated by $\pm 90^\circ$ about a vertical axis to visualize the internal cavity surfaces. Residues that contact the bound ReCl_6^{2-} ion are labeled. (B) Hydrophobic surface representation colored as in Figure 2B. (C) Complexation of ReCl_6^{2-} within the central cavity. Interacting side chains of CadC within a distance of 4 Å to the rhenate central ion are indicated. (D) Results from a docking simulation with cadaverine. The three cadaverine conformations with highest scores as well as the putative interacting residues in the cavity are depicted as sticks. Cadaverine carbon atoms are colored differently according to scoring rank order: 1, yellow; 2, salmon; 3, green. Oxygen atoms are shown in red, nitrogen atoms are shown in blue, hydrogen atoms are shown in white, and hydrogen bonds are indicated as dark green dashed lines.

Figure 4. Comparison of the crystal structure and secondary structure topology of CadC_{pd} with its most closely related α -helical bundle proteins. Subdomains to be considered for the comparison are colored whereas other structural elements are shown in gray. (A) CadC_{pd} with its C-terminal subdomain (residues 333–512) comprising 11 α -helices (salmon). (B) The peroxisomal targeting signal 1 binding domain of *Trypanosoma brucei* peroxin 5 (PDB entry 3CVN) forms a twisted 16 α -helix bundle; helices 3–12 are colored green. (C) The transcription factor MalT domain III protein of *Escherichia coli* (1HZ4) forms another twisted 18 α -helix bundle; helices 6–15 are colored blue. (D) The tetratricopeptide repeat protein Nlpl of *Escherichia coli* (1XNF) with α -helices 4–13 colored yellow.

Acidobacteria bacterium (94969422) with 27% identity to the CadC sequence 248–387. Interestingly, several CadC orthologs revealed conserved residues at the positions that line the putative cadaverine cavity described above (Fig. 5).

Discussion

CadC belongs to the ToxR-like regulators that encompass biochemically non-modified one-component systems with similar gross topology, including several low pH-induced transcription regulators.

Figure 5. Sequence alignment of CadC_{pd} orthologs (residues 188–512) from a representative set of γ-proteobacteria. Amino acids lining the characteristic cavity of CadC_{pd} are labeled as small vertical arrows above the sequences while residues involved in the dimer interface are indicated by arrows underneath, many of which involve fully conserved positions. Secondary structure elements are shown as horizontal arrows (β-strands) and rods (α-helices), respectively, above the alignment according to the crystal structure of the *Escherichia coli* CadC_{pd} protein determined in this study. Amino acid Asp471, which is important for pH-sensing, is marked by an asterisk. GI numbers (www.ncbi.nlm.nih.gov/Sitemap/sequenceIds.html) are given in parentheses: *Escherichia coli* MG1655 (16131959), *Escherichia fergusonii* ATCC 35469 (218549242), *Salmonella enterica* serovar typhimurium LT2 (16765877), *Serratia proteamaculans* 568 (157323768), *Yersinia ruckeri* ATCC 29473 (238756479), *Klebsiella pneumoniae* NTUH-K2044 (218117876), *Edwardsiella tarda* EIB202 (267983771).

Because of the important role of these proteins for bacterial virulence, for example ToxR of *Vibrio cholerae*, there is elevated interest in their biomolecular mechanism. Albeit there is no clear correlation between the primary structures of the periplasmic domains of CadC and ToxR, even in terms of size (CadC_{pd}: 325 amino acids; ToxR_{pd}: 100 amino acids), it is conceivable that the general mechanism of signal perception and transduction across the inner bacterial membrane should be related.

The X-ray analysis of CadC_{pd} reported here provides the first known tertiary structure for a sensory domain of a (pH-activated) ToxR-like regulator. A characteristic feature of the CadC_{pd} structure is its composition of two subdomains with a cavity at their interface that is suited to accommodate cadaverine, the feedback inhibitor of the Cad system. The docking results demonstrate that the side chains of the residues Glu447, Asp225, Thr229, Tyr374, and Gln421 within this pocket are appropriately oriented to bind the positively charged ligand. This finding is in agreement with the binding of the hexachlororhenate ion in almost the same place. In this case, the side chains of residues Asp225, Thr229, His344, Glu447, and Trp450 are involved. Notably, the predicted dissociation constants for the CadC_{pd}-cadaverine complex from the docking experiment of 12–28 μM are in the same order of magnitude as the experimentally determined value of $96 \pm 18 \mu\text{M}$.¹⁴ The lower estimated value may be caused by incorrect assumptions for the electrostatic contribution to the binding interaction. However, the measured moderate affinity explains the lack of success in our soaking experiments with this ligand. Generally, the flexibility of the cadaverine ligand and the preponderance of electrostatic interactions should facilitate adaptation of the mode of binding to partly different charged and/or polar residues that line the ligand pocket in other species (cf. Fig. 5). For example, it seems that in bacteria that carry Gly instead of Asp at position 225 the role of the charged side chain could be taken over by the adjacent Asp229 (instead of Thr).

Another remarkable feature of the CadC_{pd} crystal structure is the formation of a biological dimer as revealed by PISA and CS analysis and experimentally backed by size exclusion chromatography. The quaternary structure of this dimer would allow an arrangement with respect to the lipid bilayer that is compatible with a preceding transmembrane segment leading into the N-terminal subdomain dominated by the five-stranded β -sheet. Thus, it is conceivable that pH-dependent protonation of negatively charged side chains, including residue Asp471, should affect the interface of the dimer, resulting in an altered mutual orientation of the periplasmic domains and, thus, to a general conformational effect that becomes transduced via the transmembrane segment and ultimately triggers

transcriptional activity on the cytoplasmic side. The observation that the mutation D471N (representing the ON state) both has a pronounced effect on the solubility of the protein (see Materials and Methods) and leads to a different crystal packing supports the notion of minute conformational changes at the dimer interface that should functionally influence transcriptional activity. This would also be consistent with an additional role of cadaverine binding at the cavity within each periplasmic domain on the mode of dimer association, apart from the pH-sensing effect. Notably, the complex between CadC_{pd} and the rhenate ligand showed a significantly smaller interface area of 1368 $\text{\AA}^2$ per monomer with fewer hydrogen bonds (4) and salt bridges (2).

In conclusion, the three-dimensional structure of CadC_{pd} elucidated in this study establishes a new protein superfamily as there is no significant structural similarity to any other protein in the PDB. Only sub-structures of three large α -helical bundle proteins show certain resemblance to the C-terminal α -helical subdomain of CadC_{pd}. In particular, there is no structurally related protein with the N-terminal β -sheet subdomain of CadC_{pd}. The results from a BLAST search underline these findings because there is no independent protein in the database that shows significant sequence similarity to CadC_{pd}, including members of the ToxR-like protein family itself. However, CadC homologues with high sequence similarity are found in several γ -proteobacterial genera like the *Salmonella*, *Serratia*, and *Klebsiella*. Thus, it can be expected that these CadC-related proteins form a novel structural family with similar functions in bacterial cell physiology. Mutagenesis experiments are currently under way to further study the structural mechanisms that underly the pH- and ligand-dependent transcriptional activation of this interesting bacterial membrane sensor protein.

Materials and Methods

Overexpression and purification of wild type CadC_{pd} and its variants D471E and D471N

CadC_{pd} was produced as a fusion protein consisting of N-terminal thioredoxin, a double His₆-tag interspersed with thrombin and enterokinase cleavage sites, and the CadC fragment comprising residues 188–512 (Swiss-Prot entry P23890). This hybrid protein was overproduced in *Escherichia coli* Origami B (DE3)/pLysS using the plasmid pET32a-CadC188-512 as described before.¹⁴ The structural genes for the CadC_{pd} variants D471E and D471N were obtained by site-directed mutagenesis and expressed like the wild type protein.¹⁷ Briefly, the fusion proteins were purified from the bacterial whole cell extract by Ni-NTA affinity chromatography. After that, thioredoxin was cleaved off using the Thrombin Cleavage Capture Kit

(Merck, Darmstadt, Germany). The reaction product was dialyzed against 20 mM Tris/HCl pH 8.0, 130 mM NaCl and applied to a Q Sepharose Fast Flow anion exchange column (GE Healthcare, Munich, Germany), followed by elution with a gradient ranging from 130 mM to 1 M NaCl in 20 mM Tris/HCl pH 8.0. Fractions containing CadC_{pd} were pooled and concentrated to 11 mg/mL by ultrafiltration. The wild type protein was finally purified by size exclusion chromatography using an S75 analytic HR 10/30 column (GE Healthcare) in the presence of 10 mM Tris/HCl pH 7.5, 150 mM NaCl. CadC_{pd} was then concentrated with 10 kDa cut-off Vivaspin tubes (Vivascience, Hannover, Germany) yielding two batches with concentrations of 8.8 and 13 mg/mL, respectively, and filtrated with 0.45 µm Spin-X centrifuge tubes (Corning Inc., Corning, NY) before crystallization. To measure the apparent size of the wild type CadC_{pd} protein, a size exclusion chromatography was performed on a calibrated S200 analytic HR 10/30 column (GE Healthcare) in the presence of 150 mM NaCl, 10 mM Tris/HCl pH 7.5. The CadC_{pd} variant D471E was purified in the same way and finally concentrated to 14 mg/mL. The variant D471N, however, formed a gel-like precipitate after the first dialysis against 20 mM Tris/HCl pH 8.0, 130 mM NaCl, resulting in an initial protein concentration of merely 2.8 mg/mL. Following centrifugation and sterile filtration, the protein was purified via anion exchange chromatography as described above, and, after adjusting the NaCl concentration to 500 mM, subjected to ultrafiltration. The concentrated sample, which turned opaque after resting on ice for an hour, was cleared by filtration through a Spin-X tube, resulting in a final protein concentration of 6.9 mg/mL, and used for crystallization trials without size exclusion chromatography.

Protein crystallization and preparation of heavy atom derivatives

CadC_{pd} was crystallized using the hanging drop vapor diffusion technique, yielding crystals within 1 week at 20°C. The precipitants were 1.4 to 1.7 M (NH₄)₂SO₄ in combination with 4.5–11% (v/v) 2-propanol (without additional buffer except for the 10 mM Tris/HCl pH 7.5 from the original protein solution; cf. above). Crystals of native CadC_{pd} were harvested from their drops, dipped into the reservoir solution supplemented with 30% (v/v) glycerol, and flash-frozen. Crystals of the variant D471E were grown under the same conditions, whereas crystals of the variant D471N were obtained in an acidic milieu with 20% (w/v) PEG 6000, 100 mM Na-citrate pH 5. In the latter case, the droplets were set up with a 2:1 protein to precipitant volume ratio to compensate for the lower concentration of the original protein solution. A search for heavy atom compounds (Hampton Research, Aliso Viejo, CA) that bind to wild type

CadC_{pd} was performed by means of native polyacrylamide gel electrophoresis (PAGE).²⁹ To this end, 5 µL of each heavy atom compound with a concentration of 10 mM in aqueous solution was added to 5 µL of a 2 mg/mL CadC_{pd} solution in 10 mM Tris/HCl pH 7.5, 150 mM NaCl and the mixture was applied to 6.5% (w/v) PAGE in the presence of 750 mM Tris/HCl pH 7.5 using 190 mM glycine, 10 mM Tris/HCl pH 8.3 as running buffer. A detectable band shift of the protein-heavy atom complex led to the identification of K₂ReCl₆ as a promising compound. To prepare the heavy atom derivative for X-ray data collection, crystals of CadC_{pd} were first transferred from the mother liquor described above to 2.5 M Li₂SO₄ to avoid the formation of metal-ammonia complexes. Then, a solution of 30 mM K₂ReCl₆ and 1.7 M Li₂SO₄ was cautiously added to the drop such that the final ReCl₆²⁻ concentration reached 5 mM. Within about 10 min the initially pale yellow color of the droplet changed to dark violet. On the next day, the deeply colored crystals were washed in 2.5 M Li₂SO₄ to remove precipitated rhenium salt. Then, the soaking was repeated for 4 h, again followed by washing of the crystals with 2.5 M Li₂SO₄. Finally, crystals were transferred to 2.5 M Li₂SO₄, 30% (v/v) glycerol, using MicroMount loops (MiTeGen, Ithaca, NY) and flash-frozen in liquid nitrogen.

Data collection and processing, model building, and refinement

For the structure solution of CadC_{pd} via MAD, three datasets from a single Re-soaked crystal of space group P6₁22 were collected at BESSY beam line 14.2. Diffraction up to a resolution of 2.3 Å (peak), 2.6 Å (inflection), and 2.9 Å (high energy remote) was measured at 100 K (Table I). Data were processed with MOSFLM, scaled with SCALA,³⁰ and heavy atom sites were identified with SHELXD.³¹ Refinement of the Re sites, MAD phasing, solvent flattening, and calculation of a native electron density map were carried out with SHARP.³² The initial atomic model of the Re-complexed CadC_{pd} was built using ARP/wARP³⁰ and manually corrected with Coot.³³ Altogether 16 ReCl₆²⁻ sites were verified by inspecting the anomalous density map, and their occupancy was adjusted to yield B-factors similar to surrounding protein atoms. The geometry of the ReCl₆²⁻ complex ion itself was obtained from the Inorganic Crystal Structure Database,³⁴ revealing Re-Cl distances in the range from 2.353 to 2.367 Å, consistent with literature data.³⁵ Hence, restraints during refinement were set for Re-Cl distances to 2.36 Å and for Cl-Re-Cl angles to 90°. After initial refinement of the protein-rhenate complex the crystal structures of the native protein and its variants were solved by molecular replacement with Phaser³⁶ using corresponding data sets collected at BESSY beam lines 14.1 and 14.2 (Table I). In the case of the

D471N crystal, diffraction data were processed with XDS.³⁷ Water molecules were added to all models using ARP/wARP, rotamers of Asn and Gln residues were adjusted with NQ-Flipper³⁸ and, subsequently, all protein models were refined with Refmac5³⁰ in iterative cycles interspersed with manual correction.

The electron density map of native CadC_{pd} at a resolution of 1.8 Å (Table I) allowed model building for 314 of the 362 residues present in the recombinant protein, namely Ser190–Lys206 and Tyr215–Ser511. For the ReCl₆²⁻ derivative of CadC_{pd}, there was electron density for altogether 286 residues: Ile192–Val204, Leu217–Tyr245, Ser251–Phe261, Phe273–Val293, and Asn298–Leu509. For the CadC_{pd} variant D471E the electron density map at 1.9 Å resolution covered residues Lys189–Lys206 and Tyr215–Ser511, while in the case of D471N there was electron density at 2.2 Å resolution for residues Ser190–Val204, Gln216–Asn263, Cys272–Glu394, and Ala399–Ser511. The C-terminal residue Glu512 as well as the N-terminal His₆-tag, including the remaining upstream linker of the processed fusion protein (47 residues), were invisible in all structures. The models were validated with PROCHECK³⁹ and WHAT_CHECK.⁴⁰ Secondary structure elements were assigned using DSSP⁴¹ and secondary structure topologies were depicted with TopDraw.⁴² Crystal packing contacts and protein interfaces were analyzed with PISA¹⁹ and with SC²⁰ from CCP4.³⁰ Graphics were prepared with PyMOL⁴³ and GRASP.⁴⁴ Structural superposition was carried out with PyMOL while docking studies were performed with AutoDock 4.²¹ Evolutionary related CadC_{pd} sequences were searched with BLAST²⁷ in the NCBI non-redundant protein sequences database.²⁸ Sequence alignments were subsequently made with ClustalW2⁴⁵ and edited using Aline⁴⁶ with a low similarity cut off value of 0.241 (similarity groups DE FWY HKR ILMV NQ ST). Structurally related proteins were searched with DALI²² and HHPred.²⁵ The atomic coordinates and structure factors of the CadC_{pd} models have been deposited at the RCSB Protein Data Bank under the accession codes [3LY7](#) (wild type protein), [3LYA](#) (Re-derivative), [3LY8](#) (variant D471E), and [3LY9](#) (variant D471N).

Acknowledgment

The authors wish to thank Korinna Burdack and Josef Danzer for technical assistance and Eberhardt Herdtweck and Martin U. Schmidt for database searches. The synchrotron measurements at BESSY and Free University Berlin at BESSY beamlines 14.1 and 14.2 were supported by the BMBF (project funding reference number 05 ES3XBA/5); we thank Jörg Schulze, Georg Zocher, and Sandra Pühringer for assistance. I.H. received a research scholarship from the Elite Network of Bavaria.

References

1. Padan E, Zilberstein D, Rottenberg H (1976) The proton electrochemical gradient in *Escherichia coli* cells. *Eur J Biochem* 63:533–541.
2. Gale EF, Epps HM (1942) The effect of the pH of the medium during growth on the enzymic activities of bacteria (*Escherichia coli* and *Micrococcus lysodeikticus*) and the biological significance of the changes produced. *Biochem J* 36:600–618.
3. Sokawatmaekhin W, Kuraishi A, Sakata K, Kashiwagi K, Igarashi K (2004) Excretion and uptake of cadaverine by CadB and its physiological functions in *Escherichia coli*. *Mol Microbiol* 51:1401–1412.
4. Watson N, Dunyak DS, Rosey EL, Slonczewski JL, Olson ER (1992) Identification of elements involved in transcriptional regulation of the *Escherichia coli* cad operon by external pH. *J Bacteriol* 174:530–540.
5. Merrell DS, Camilli A (2000) Regulation of *Vibrio cholerae* genes required for acid tolerance by a member of the “ToxR-like” family of transcriptional regulators. *J Bacteriol* 182:5342–5350.
6. Miller VL, Taylor RK, Mekalanos JJ (1987) Cholera toxin transcriptional activator *toxR* is a transmembrane DNA binding protein. *Cell* 48:271–279.
7. Häse CC, Mekalanos JJ (1998) TcpP protein is a positive regulator of virulence gene expression in *Vibrio cholerae*. *Proc Natl Acad Sci USA* 95:730–734.
8. Yang Y, Isberg RR (1997) Transcriptional regulation of the *Yersinia pseudotuberculosis* pH6 antigen adhesin by two envelope-associated components. *Mol Microbiol* 24:499–510.
9. Egan S, James S, Kjelleberg S (2002) Identification and characterization of a putative transcriptional regulator controlling the expression of fouling inhibitors in *Pseudoalteromonas tunicata*. *Appl Environ Microbiol* 68:372–378.
10. Seki T, Yoshikawa H, Takahashi H, Saito H (1987) Cloning and nucleotide sequence of *phoP*, the regulatory gene for alkaline phosphatase and phosphodiesterase in *Bacillus subtilis*. *J Bacteriol* 169:2913–2916.
11. Melchers LS, Thompson DV, Idler KB, Schilperoort RA, Hooykaas PJ (1986) Nucleotide sequence of the virulence gene *virG* of the *Agrobacterium tumefaciens* octopine Ti plasmid: significant homology between *virG* and the regulatory genes *ompR*, *phoB* and *dye* of *E. coli*. *Nucleic Acids Res* 14:9933–9942.
12. Comeau DE, Ikenaka K, Tsung KL, Inouye M (1985) Primary characterization of the protein products of the *Escherichia coli* *ompB* locus: structure and regulation of synthesis of the OmpR and EnvZ proteins. *J Bacteriol* 164:578–584.
13. Kuper C, Jung K (2005) CadC-mediated activation of the *cadBA* promoter in *Escherichia coli*. *J Mol Microbiol Biotechnol* 10:26–39.
14. Tetsch L, Koller C, Haneburger I, Jung K (2008) The membrane-integrated transcriptional activator CadC of *Escherichia coli* senses lysine indirectly via the interaction with the lysine permease LysP. *Mol Microbiol* 67:570–583.
15. Tetsch L, Jung K (2009) The regulatory interplay between membrane-integrated sensors and transport proteins in bacteria. *Mol Microbiol* 73:982–991.
16. Dell CL, Neely MN, Olson ER (1994) Altered pH and lysine signalling mutants of *cadC*, a gene encoding a membrane-bound transcriptional activator of the *Escherichia coli* *cadBA* operon. *Mol Microbiol* 14:7–16.
17. Haneburger I, Eichinger A, Skerra A, Jung K (2011) New insights into the signaling mechanism of the pH-

- responsive, membrane-integrated transcriptional activator CadC of *Escherichia coli* J Biol Chem, in press. [doi:10.1074/jbc.M110.196923].
18. Fritz G, Koller C, Burdack K, Tetsch L, Haneburger I, Jung K, Gerland U (2009) Induction kinetics of a conditional pH stress response system in *Escherichia coli*. J Mol Biol 393:272–286.
 19. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372:774–797.
 20. Lawrence MC, Colman PM (1993) Shape complementarity at protein/protein interfaces. J Mol Biol 234:946–950.
 21. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comp Chem 19:1639–1662.
 22. Holm L, Kaariainen S, Rosenstrom P, Schenkel A (2008) Searching protein structure databases with DaliLite v.3. Bioinformatics 24:2780–2781.
 23. Sampathkumar P, Roach C, Michels PA, Hol WG (2008) Structural insights into the recognition of peroxisomal targeting signal 1 by *Trypanosoma brucei* peroxin 5. J Mol Biol 381:867–880.
 24. Steegborn C, Danot O, Huber R, Clausen T (2001) Crystal structure of transcription factor MalT domain III: a novel helix repeat fold implicated in regulated oligomerization. Structure 9:1051–1060.
 25. Söding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33:W244–248.
 26. Wilson CG, Kajander T, Regan L (2005) The crystal structure of NlpI. A prokaryotic tetratricopeptide repeat protein with a globular fold. FEBS J 272:166–179.
 27. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410.
 28. Bleasby AJ, Akrigg D, Attwood TK (1994) OWL—a non-redundant composite protein sequence database. Nucleic Acids Res 22:3574–3577.
 29. Boggon TJ, Shapiro L (2000) Screening for phasing atoms in protein crystallography. Structure 8:R143–149.
 30. CCP4 (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 50:760–763.
 31. Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. Acta Crystallogr D Biol Crystallogr 58:1772–1779.
 32. Bricogne G, Vonrhein C, Flensburg C, Schiltz M, Paciorek W (2003) Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta Crystallogr D Biol Crystallogr 59:2023–2030.
 33. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60:2126–2132.
 34. Bergerhoff G, Brown ID, Inorganic Crystal Structure Database. In Allen FH, Bergerhoff G, Sievers R, Eds. (1987) Crystallographic Databases. Chester: IUCr.
 35. Barbour LJ, MacGillivray LR, Atwood JL (1996) Crystal and molecular structure of [H₃O•18-crown-6]₂[ReCl₆] isolated from a liquid clathrate medium. J Chem Cryst 26:59–61.
 36. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. J Appl Cryst 40:658–674.
 37. Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J Appl Cryst 26:795–800.
 38. Weichenberger CX, Sippl MJ (2007) NQ-Flipper: recognition and correction of erroneous asparagine and glutamine side-chain rotamers in protein structures. Nucleic Acids Res 35:W403–406.
 39. Laskowski RA, MacArthur MW, Mos DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 26:283–291.
 40. Hoof RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. Nature 381:272.
 41. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637.
 42. Bond CS (2003) TopDraw: a sketchpad for protein structure topology cartoons. Bioinformatics 19:311–312.
 43. DeLano WL (2002) The PyMOL Molecular Graphics System. California: DeLano Scientific.
 44. Nicholls A, Sharp KA, Honig B (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 11:281–296.
 45. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948.
 46. Bond CS, Schüttelkopf AW (2009) ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. Acta Crystallogr D Biol Crystallogr 65:510–512.
 47. Eichinger A, Nasreen A, Kim HJ, Skerra A (2007) Structural insight into the dual ligand specificity and mode of high density lipoprotein association of apolipoprotein D. J Biol Chem 282:31068–31075.