doi:10.1016/j.jmb.2008.10.061 J. Mol. Biol. (2009) 386, 962–975

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Crystal Structure of SpoVT, the Final Modulator of Gene Expression during Spore Development in *Bacillus subtilis*

Iris Asen[1]†, Sergej Djuranovic[2]†, Andrei N. Lupas[2] and Kornelius Zeth[2]*

[1]*Department of Membrane Biochemistry, Max Planck Institute for Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany*

[2]*Department of Protein Evolution, Max Planck Institute for Developmental Biology, Spemannstrasse 35, 72076 Tübingen, Germany*

Received 4 August 2008; received in revised form 21 October 2008; accepted 22 October 2008
Available online 30 October 2008

Edited by R. Huber

Endospore formation in *Bacillus subtilis* is orchestrated by five developmental sigma factors and further modulated by several auxiliary transcription factors. One of these, SpoVT, regulates forespore-specific σ^G-dependent genes and plays a key role in the final stages of spore formation. We have determined the crystal structure of the isolated C-terminal domain of SpoVT at 1.5 Å by experimental phasing techniques and used this model to solve the structure of the full-length SpoVT at 2.6 Å by molecular replacement. SpoVT is a tetramer that shows an overall significant distortion mediated by electrostatic interactions. Two monomers dimerize *via* the highly charged N-terminal domains to form swapped-hairpin β-barrels. These asymmetric dimers further tetramerize through the formation of mixed helix bundles between their C-terminal domains, which themselves fold as GAF (cGMP-specific and cGMP-stimulated phosphodiesterases, *Anabaena* adenylate cyclases, and *Escherichia coli* FhlA) domains. The combination of a swapped-hairpin β-barrel with a GAF domain represents a novel domain architecture in transcription factors. The occurrence of SpoVT homologs throughout *Bacilli* and *Clostridia* demonstrates the ancestral origin of this factor in sporulation.

Keywords: SpoVT; sporulation; transcription factor; AbrB; GAF domain

Introduction

Gram-positive organisms, such as *Bacillus subtilis*, adapt to unfavorable conditions by protecting their genome in robust endospores. Sporulation is a complex and tightly balanced differentiation process proceeding in seven stages. Asymmetric division of the vegetative cell generates a mother cell and a small forespore (stages I and II). After engulfment of the forespore by the mother cell (stage III), a cortex consisting of a modified form of peptidoglycan is synthesized between the membranes that separate the cells (stage IV). In stage V, a second protective layer composed of approximately 50 proteins (the coat) is built around the spore. Maturation and eventual release of the spore by lysis of the mother cell (stages VI and VII) complete the sporulation cycle.[1]

Endospore formation in *B. subtilis* is controlled by the action of five sigma factors. Further modulation of gene expression within these regulons is achieved by additional transcriptional regulators: SpoIIID (σ^E dependent), GerE (σ^K dependent), RsfA (σ^F dependent), and SpoVT (σ^G dependent).[2,3] The σ^G-controlled transcription factor SpoVT coordinates gene expression in the small forespore chamber, which is destined to become the mature dormant spore. SpoVT acts not only as a transcriptional activator but also as a repressor, affecting the expression

*Corresponding author. E-mail address: kornelius.zeth@tuebingen.mpg.de.

† I.A. and S.D. contributed equally to this work.

Present address: S. Djuranovic, Department of Molecular Biology and Genetics, Howard Hughes Medical Institute, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.

Abbreviations used: GAF, cGMP-specific and cGMP-stimulated phosphodiesterases, *Anabaena* adenylate cyclases, and *Escherichia coli* FhlA; SAD, single anomalous dispersion; AU, asymmetric unit; Se-Met, selenomethionine; PEG, polyethylene glycol.

level of 47 genes.[2,4] Although most of these gene products are of yet unassigned function, germination response receptors have been identified, which are repressed, while small acid-soluble spore proteins are induced.[2,4] The latter ones associate with the densely packed DNA in the spore for protective purposes. *spoVT* mutants have a poorly assembled spore coat and are a germination defective.[5,6] SpoVT thus plays an indispensable role in the late stage (stage V) of spore formation and hence ensures survival of the genetic information.[5,6]

SpoVT of *B. subtilis* is a protein of 178 amino acids, with a molecular mass of 19.7 kDa. It contains two domains: the N-terminal domain affords DNA binding, whereas the C-terminal is of importance for the individual function of SpoVT.[7] The DNA-binding domain has 68% sequence identity to the N-terminal region of the transcription regulator AbrB,[7] a key transition-state regulator of *B. subtilis* that acts during the transition from vegetative growth to sporulation.[8] Therefore, SpoVT belongs to the superfamily of swapped-hairpin transcription factors, which fold into homodimeric β-barrels *via* four pairs of interleaved β-hairpins.[9] The C-terminal domain of SpoVT was predicted to be a GAF (cGMP-specific and cGMP-stimulated phosphodiesterases, *Anabaena a*denylate cyclases, and *Escherichia coli F*hlA)-like domain by sequence comparisons.[10] GAF domains, which are represented in a broad range of sensory pathways, can bind linear and cyclic nucleotides, porphyrin rings, and signaling molecules, such as the autoinducer *N*-(3-oxo-octanoyl)-L-homoserine lactone.[11–14] The GAF domain of the *B. subtilis* transcription factor CodY, which is a sentinel of the nutrient state, binds GTP and the branched-chain amino acids isoleucine and valine.[15,16] Other GAF domains do not bind ligands and may have a pure structural function to mediate dimerization.[12,17] The fold of these domains is typically of compact mixed α/β architecture, with a core sheet of five or six anti-parallel β-strands flanked by α-helices on one side and by the largely accessible ligand binding pocket on the other.

The gene encoding SpoVT represents a novel fusion of an AbrB-like sequence to a GAF domain. We have

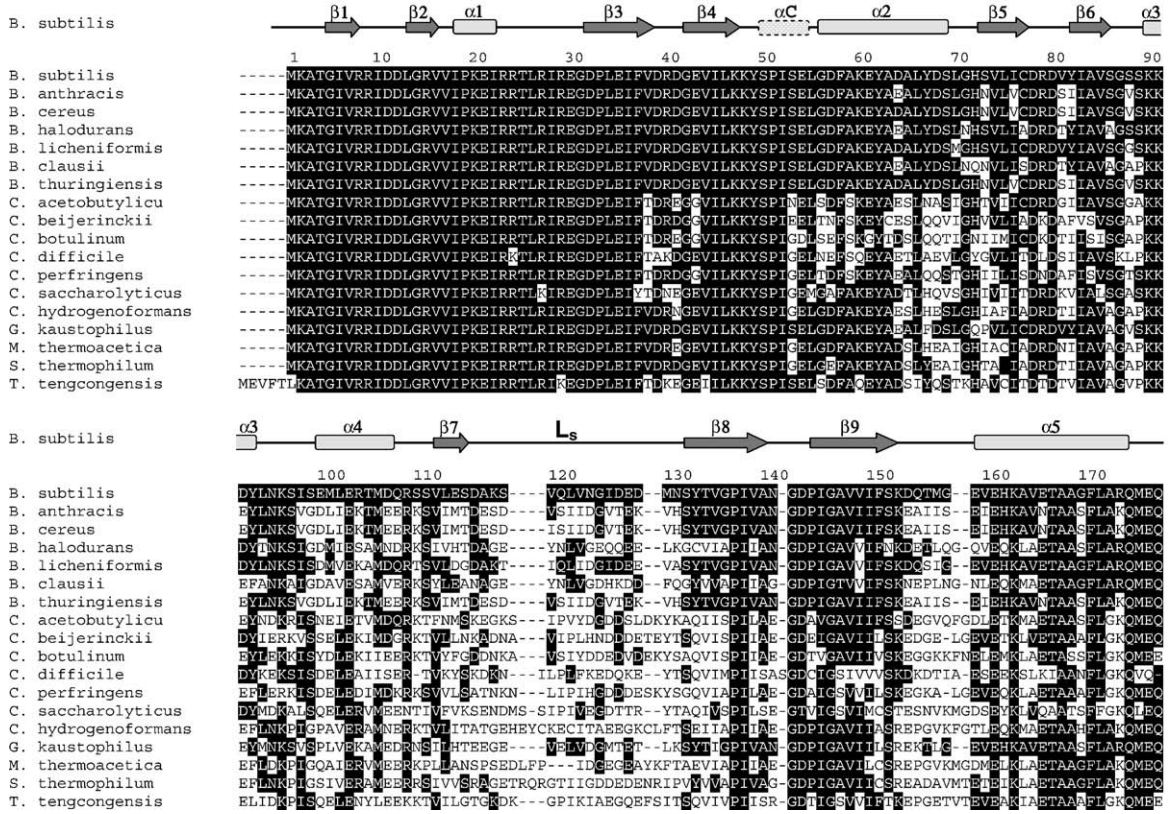

Fig. 1. Sequence Alignment of SpoVT Orthologs. *Bacillus subtilis* (*B. subtilis*); *Bacillus anthracis*; *Bacillus cereus* (*B. cereus*); *Bacillus halodurans* (*B. halodurans*); *Bacillus licheniformis* (*B. licheniformis*); *Bacillus clausii* (*B. clausii*); *Bacillus thuringiensis* (*B. thuringiensis*); *Clostridium acetobutylicum* (*C. acetobutylicum*); *Clostridium beijerinckii* (*C. beijerinckii*); *Clostridium botulinum* (*C. botulinum*); *Clostridium difficile* (*C. difficile*); CLOPS, *Clostridium perfringens* (*C. perfringens*); *Caldicellulosiruptor saccharolyticus* (*C. saccharolyticus*); *Carboxydothermus hydrogenoformans* (*C. hydrogenoformans*); *Geobacillus kaustophilus* (*G. kaustophilus*); *Moorella thermoacetica* (*M. thermoacetica*); *Symbiobacterium thermophilum* (*S. thermophilum*); *Thermoanaerobacter tengcongensis* (*T. tengcongensis*). Invariant residues are highlighted in black boxes. Secondary structure elements of *B. subtilis* SpoVT are indicated above with β-sheets as dark arrows and α-helices as light barrels. Helix αC present in the compact conformation of monomer $SpoVT_A$ is shaded and the substrate binding pocket locking loop is marked (L_S).

initially determined the crystal structure of the C-terminal and dimeric SpoVT GAF domain, which was further used for the determination of the entire complex. In this complex, two monomers dimerize *via* their N-terminal swapped-hairpin domains into dimers. These dimers associate into tetramers through helical interactions between their GAF domains as 'pairs of dimers.' Among oligomers, a significant asymmetry of the individual monomers is imposed by the formation of electrostatic interactions. This asymmetric tension may have implications in the process of DNA recognition and binding.

Results and Discussion

SpoVT in endospore-forming organisms

Sequence searches with *B. subtilis* SpoVT (gi 586883) resulted in a set of 18 unique sequences, all of which derived from endospore-forming bacteria throughout *Bacilli* and *Clostridia* (Fig. 1). From the alignment, it is visible that the N-terminal DNA-binding domain of the protein is unusually well conserved, with a sequence identity of ~95% for the first 55 residues, including the part connecting the first domain with the second. This finding suggests that the DNA-binding properties mediated by the N-terminal domain of all orthologs are strongly conserved and tuned for highly specific promoter recognition. By contrast, the C-terminal part (residues 56–178) of the protein is much less conserved with sequence identities of about 20%–30% between two sequences (Fig. 1). Residues conserved in this domain can be roughly grouped into (1) residues stabilizing the intramolecular domain interface (e.g., Glu103; Fig. 2c), (2) amino acids stabilizing the intermolecular interface (e.g., Lys163; Fig. 2b), (3) residues embedded in fold integrity (glycines and prolines), and (4) residues stabilizing the hydrophobic core structure of the GAF domain. Interestingly, the long loop of the "lid-like" structure L_S (comprising residues 116–131), closing the potential substrate pocket (see Fig. 1 and the succeeding paragraphs), shows no conserved sequence feature and varies in length in other homologs. This finding, together with biochemical data subsequently presented, may imply a rather unspecific (if any) substrate spectrum of SpoVT homologs.

SpoVT structure determination

In order to understand the mechanism by which SpoVT and AbrB from *B. subtilis*, albeit being almost identical in sequence, can specifically discriminate DNA promoter sequences, we set out to crystallize the protein complex and determined the structure by X-ray crystallographic methods (for details, see Materials and Methods). Initially, we analyzed the structure of the isolated SpoVT GAF domain and solved this structure by single anomalous dispersion (SAD) at 2.1 Å. The resulting initial model was refined against high-resolution data of the wild-type protein collected at 1.5 Å to R_{cryst} and R_{free} values of 19.7 and 22.6 with two domains in the asymmetric unit (AU), respectively (Table 1). The structure of the full-length protein (SpoVT) was later determined at 2.6 Å by molecular replacement using the model of the isolated GAF domain and refined to R_{cryst} and R_{free} factors of 22.5 and 26.6, respectively (Table 1). SpoVT with two molecules in the AU has a Matthew coefficient of 3.2 (solvent content of 61%).[18] Details of the crystallographic data and statistics are listed in Table 1.

Structure analysis of the monomeric transcription factor SpoVT

Transcription factors often express an unusual distribution of charges due to their specific function in DNA binding. This peculiarity is also reflected in the N-terminal domain of SpoVT, which has a positive net charge to allow protein–DNA interactions (Fig. 2d). However, the entire protein carries a significant surplus of negative charges, resulting in a theoretical p*I* of 4.9, an overall charge of −8, and a distribution of every third residue being charged (Fig. 2d). A subset of these conserved residues is involved not only in intramolecular salt bridges (e.g., Arg27, Arg29, and Glu103; Fig. 2c and d), H bonds (helix formation of αC; Fig. 1) but also in salt bridges between monomers to stabilize the tetrameric ring structure (Arg40, Glu35, and Glu54; Figs. 1 and 6c and d).

The overall monomeric structure of SpoVT consists of the successive arrangement of an AbrB-like domain and a GAF domain. In the crystal structure, the monomeric SpoVT reveals an elongated shape of a mixed α/β fold with a progression of secondary structure elements as follows: β1/β2/α1/β3/β4 for the AbrB-like domain ($SpoVT_N$) and α2/β5/β6/α3/α4/β7/β8/β9/α5 for the GAF domain ($SpoVT_C$; Fig. 2a). In the monomer, the two domains are connected by a linker (residues 49–55) that represents a flexible part of the protein structure allowing mobility and potential shifting of the domains relative to each other. A small helix termed αC (Figs. 1 and 2a) in the linker region between the two domains is ordered only in one of the two protein conformers building the AU (Fig. 2e). The formation of this helix is supported by intradomain interactions through electrostatic interactions leading to a more compact structure in monomer A ($SpoVT_A$; Fig. 2c and e), while in the second monomer, the N-terminal domain is not connected to the GAF domain by side-chain interactions ($SpoVT_B$). Accordingly, the interfaces between the two domains differ substantially (Fig. 2e) and allow only monomer A formation of salt bridges between Arg27/Arg29 and Glu103 of the two domains (Fig. 2c). The residues composing the linker in $SpoVT_B$ have high *B*-factors while in the case of $SpoVT_A$ are structurally well defined and involved in interactions between opposing $SpoVT_A$ monomers in the tetramer that constrain the

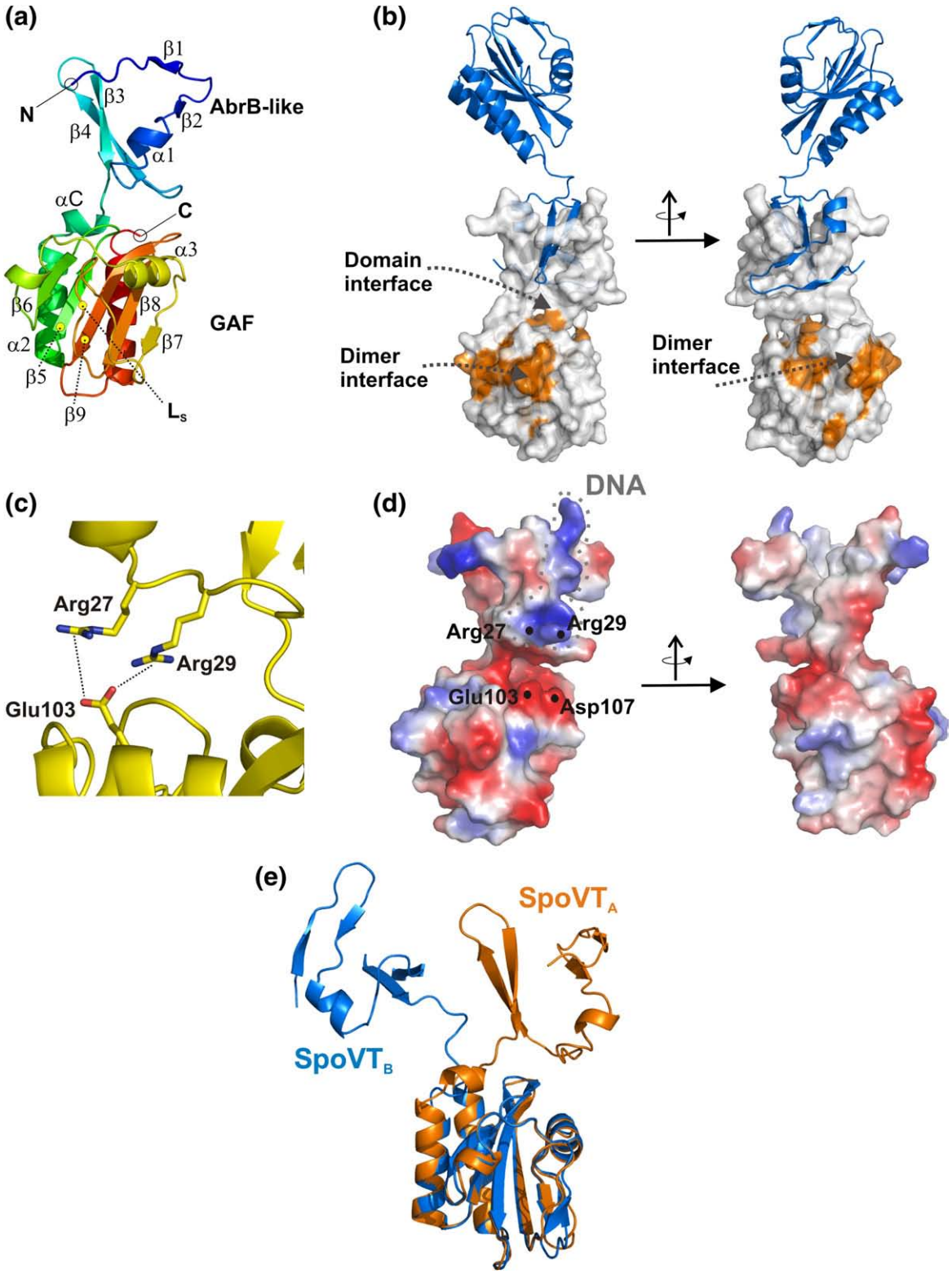

Fig. 2. Domain arrangement of monomeric SpoVT. (a) Overall structure of monomeric SpoVT with the individual domains (AbrB-like and GAF) and the secondary structure assigned (β1–β9, α1–α5, αC, and L_S). (b) Surface representation of SpoVT and assignment of conserved residues (in orange) of the GAF domain based on the alignment shown in Fig. 1. The second molecule of the AU is shown in blue. (c) Intramolecular interactions within the AbrB-like and GAF domains of the $SpoVT_A$ monomer. (d) Surface charge distribution of charged amino acids: blue for basicity and red for acidity, respectively. Several residues involved in intradomain interactions are indicated by black dots and labels. The region of putative DNA binding is surrounded by blue circles. (e) Superposition of GAF domains of the asymmetric SpoVT monomers (blue, $SpoVT_B$; orange, $SpoVT_A$).

Table 1. X-ray data collection and refinement statistics

	Native $SpoVT_{CT}$ (residues 56–178)	Se-Met $SpoVT_{CT}$ (residues 56–178)	Native $SpoVT_{FL}$ (residues 1–178)
Data collection			
Wavelength (Å)	0.979	0.979	0.978
Resolution range (Å) (outer shell)	20.00–1.50 (1.71–1.50)	20.00–2.06 (2.19–2.06)	20.00–2.60 (2.80–2.60)
Crystal form	Tetragonal	Tetragonal	Tetragonal
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$
Cell dimensions (Å)	$a=b=63.9$, $c=77.3$	$a=b=64.6$, $c=78.6$	$a=b=94.5$, $c=110.8$
Unique reflections	25,488 (3818)	19,148 (2866)	18,815 (2783)
Observed reflections	182,057 (28,670)	220,814 (31,146)	177,074 (20,904)
Redundancy	7.1 (7.5)	11.5 (10.9)	9.4 (7.5)
Completeness (%)	99.0 (100)	97.8 (91.3)	99.4 (90.4)
R_{merge}[a] (%)	6.3 (25.4)	8.5 (26.9)	10.5 (46.2)
$I/\sigma(I)$	14.8 (4.5)	14.3 (4.7)	13.9 (3.0)
Refinement			
Resolution range (Å)	20.00–1.50		20.00–2.60
Unique reflections	24,816		15,122
R_{cryst}[b]/R_{free} (%)	19.7/22.6		22.5/26.6
No. of protein atoms	1038		2708
No. of water molecules	98		9
r.m.s.d. of bond lengths (Å)	0.01		0.01
r.m.s.d. of bond angles (°)	1.1		1.0
Average *B*-factor (Å^2)			
Ramachandran plot statistics, residues (%)			
Residues in preferred regions	93.6		91.6
Residues in allowed regions	4.6		4.1
Outliers	1.8		4.4

[a] $R_{merge}=\sum_{hkl}\sum_i|I_i-\langle I\rangle|/\sum_{hkl}\sum_i\langle I\rangle$, where I_i is the intensity of the ith measurement of a reflection with indexes hkl and $\langle I\rangle$ is the statistically weighted average reflection intensity.
[b] $R_{cryst}=\sum||F_o|-|F_c||/\sum|F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively.

flexibility of the linker and lead to the formation of an α-helical turn (αC) in this region.

The N-terminal domain of SpoVT is a swapped-hairpin β-barrel

The N-terminal domain of SpoVT is a homodimeric eight-stranded swapped-hairpin β-barrel formed by the interleaving of two β-hairpins from each monomer. The intertwined architecture of the β-barrel is illustrated in Fig. 3a with a segment topology of the entire barrel of β1/β2/β2′/β1′/β3/β4/β4′/β3. The extensive dimerization interface of 1670 Å^2[19] is stabilized by a network of hydrogen bonds (16 interchain, 5 intrachain). The β-barrel is capped on both ends by the symmetry-related helices α1 and α1′ (Fig. 3a). However, the asymmetry of the two monomers seen in the complete structure is imposed onto the β-barrel geometry which does not exactly follow the twofold symmetry and leads to a partial opening between strand β1′ and β2′ of the $SpoVT_B$ monomer and a local breakdown of the continous H-bond pattern.

Due to a surplus of arginines and lysines (four lysines and eight arginines per AbrB-like monomer), $SpoVT_N$ has a higher charge density with a net positive charge of 6 per dimer, particularly in the α1-helices that contribute to a girdle of positive charges surrounding the protruding β1–β2 and β1′–β2′ loops (Fig. 3b). Most of the arginines are arranged along one face of the swapped-hairpin β-barrel and may generate a locally high charge distribution (Fig. 3b), a feature that is also known for the closely related transcription factor AbrB (68% sequence identity).[8,20]

A superposition of $SpoVT_N$ and the N-terminal part of AbrB [$AbrB_N$, Protein Data Bank (PDB) entry 1YSF] yields a root mean square deviation (r.m.s.d.) of 1 Å over 48 superimposed C^α atoms,[21] indicating structural similarity (Fig. 3c). Arg10, Arg17, Arg25, and Arg26 in AbrB are essential for DNA-binding activity[8,20] and are conserved between the two swapped-hairpin β-barrels[5,6] and in SpoVT homologs (Arg8, Arg15, Arg23, and Arg24 in SpoVT; Fig. 3c). The arginines in the anti-parallel β2/β2′-sheet (Arg15; Fig. 3c) are suitably oriented for docking into the major groove of B-DNA, as tested for AbrB.[8,20] Further arginines (Arg23 and Arg24) located in the capping α-helices support the DNA-binding interactions, as proposed for AbrB.[22] Similar charge properties are observed in another swapped-hairpin transcription factor, MazE from *E. coli*,[23,24] which acts as the antidote of a toxin–antitoxin system. Two additional arginines in $SpoVT_N$, Arg27 and Arg29 (Fig. 3c), at the extension are not present in AbrB and may putatively participate in interactions with DNA and thereby contribute to the DNA-binding specificity of $SpoVT_N$.

The C-terminal domain of SpoVT is a GAF domain

The C-proximal domain of SpoVT is of the GAF type, which is composed of the central five-stranded anti-parallel β-sheet with the strand order β6–β5–

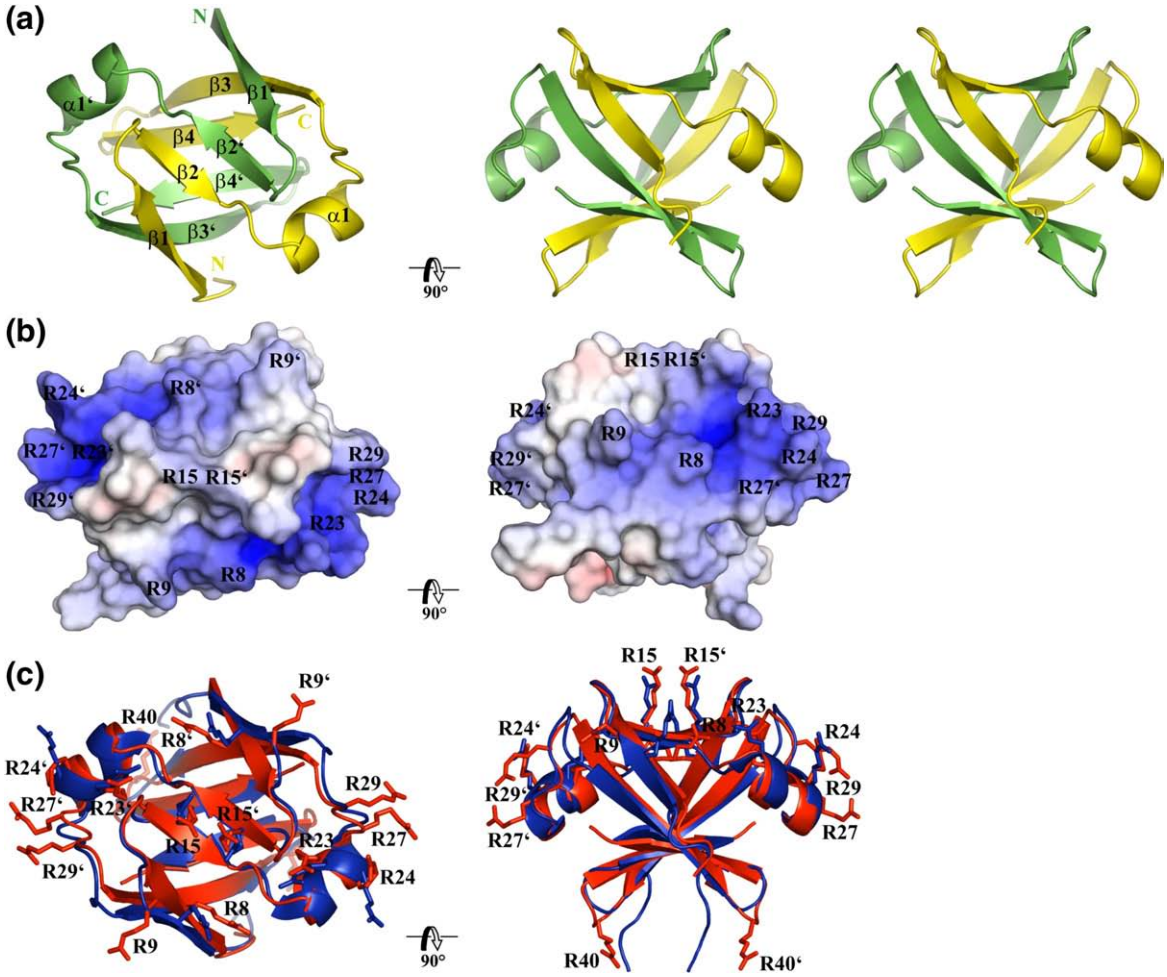

Fig. 3. N-terminal domain of SpoVT ($SpoVT_N$). (a) Overall structure of $SpoVT_N$ dimer with monomers shown in green and yellow. β-Barrel in top view with sheets, helices, and N- and C-termini labeled (left) and in side view as a stereogram (right). (b) Surface charge potential of the $SpoVT_N$ β-barrel illustrated in a surface representation with domain orientations according to (a). Charged amino acids are shown in blue and red for basicity and acidity, respectively (coloring, −10 to 10 kT/e). Arginines, except Arg40, are indicated. (c) Top (left) and side (right) views of superposed swapped-hairpin β-barrels of $SpoVT_N$ (red) and $AbrB_N$ (blue). Arginines are shown as sticks and for $SpoVT_N$ with labels in addition.

β9–β8–β7 and flanked by α-helix pairs on both sides (Fig. 4a). On one side, helices α2 and α5 are arranged as a two-helix bundle, while on the opposing side, the shorter helices α3 and α4 form, together with the extended β7–β8 loop L_S, the putative ligand binding pocket.

In sequence comparisons with protein structures, $SpoVT_C$ is most similar to the GAF domain of *B. subtilis* CodY ($CodY_N$, PDB entry 2B18), with a sequence identity of 14% for 123 residues as shown in Fig. 4b, and to two GAF domains of proteins of unknown function: *E. coli* YebR (1VHM) and *Geobacter sulfurreducens* GSU1429 (2QYB). Structure comparisons return best matches to *Thermotoga maritima* IclR (1MKM) with a *Z* score of 13.2 (r.m.s.d. of 1.5 Å over 101 superimposed C^α atoms) and to $CodY_N$ with a *Z* score of 11.0 (r.m.s.d. of 1.5 Å over 94 superimposed C^α atoms; Fig. 4c).[21] The areas of sequence and/or structure conservation between the related $SpoVT_C$ and $CodY_N$ GAF domains are restricted to secondary structure elements.

GAF domains frequently bind small molecules for regulatory or sensory purposes. A broad spectrum of ligands for GAF domains implies structural variability in their predestined binding sites (Fig. 4c). Most ligands are deeply embedded[13,25] and enclosed by loops and helices *via* a "clamp mechanism."[26] *T. maritima* IclR binds a Zn atom (Fig. 4c),[26] TraR from *Agrobacterium tumefaciens* locks *N*-(3-oxo-octanoyl)-L-homoserine lactone,[13] and the GAF-A and GAF-B domains from the product of the *Anabaena* adenylyl cyclase *cyaB2* gene bind negatively charged molecules.[25] In these cases, the binding pocket is partially charged, reflecting the polar nature of the ligand. The $SpoVT_C$ binding pocket, however, is lined by mostly hydrophobic residues (Leu75, Ile98, Leu102, Val120, Leu122, Val123, Val135, Val149, and Phe151; Fig. 4d), suggesting a hydrophobic ligand. The related $CodY_N$ in its holo form has bound hydrophobic ligands, isoleucine (PDB entry 2B18; Fig. 4c) and valine (PDB entry 2HGV), with the branched-chain

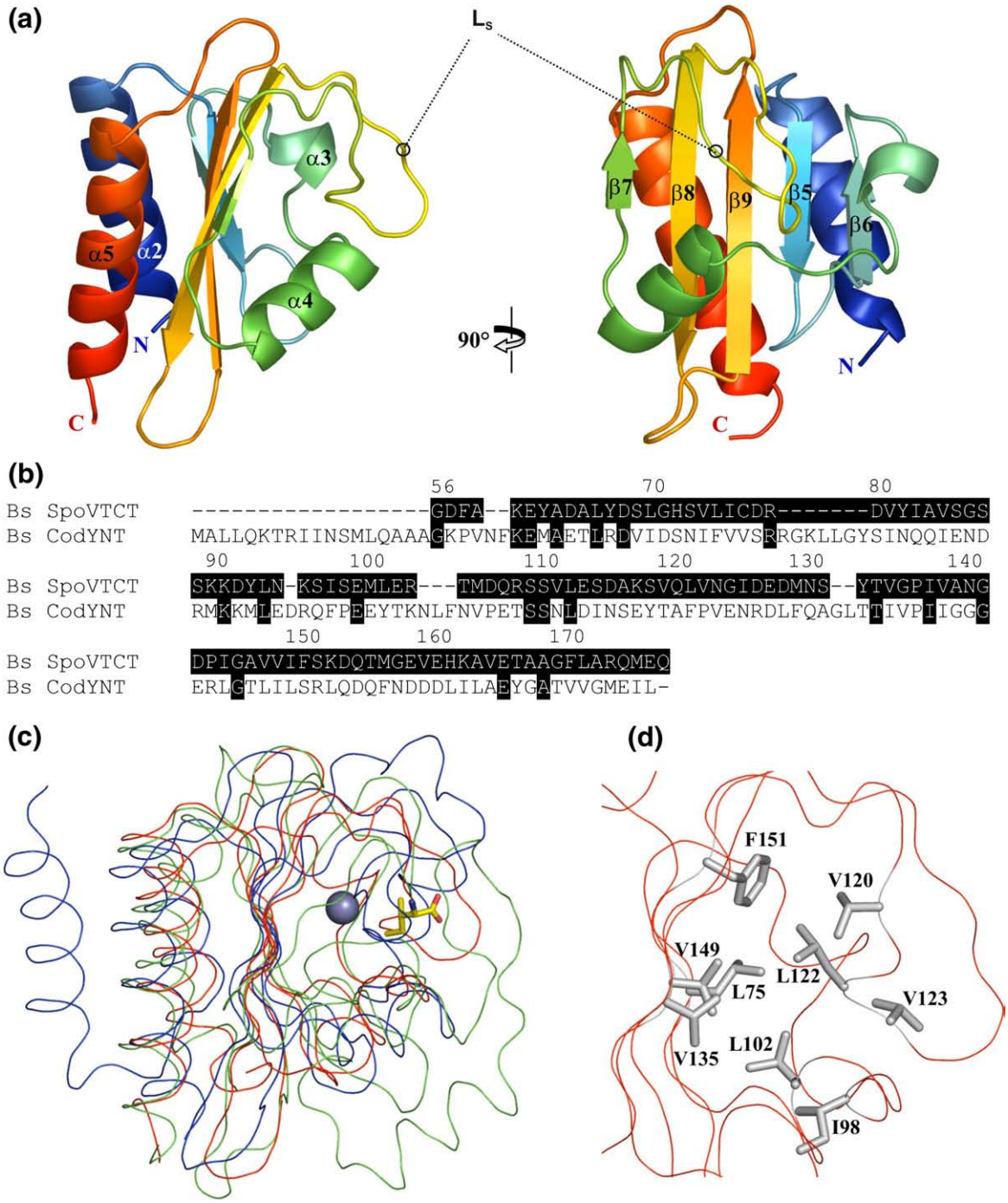

Fig. 4. C-terminal domain of SpoVT ($SpoVT_C$). (a) Overall fold of $SpoVT_C$ in rainbow color code with focus at the helices (left) and at the core five-stranded anti-parallel β-sheet (right). Secondary structure elements are labeled accordingly. (b) Sequence alignment of *B. subtilis* $SpoVT_C$ (residues 56–178, Bs $SpoVT_{CT}$) and *B. subtilis* $CodY_N$ (residues 1-155, Bs $CodY_{NT}$). Invariant residues are highlighted in black boxes. (c) Ribbon representation of superimposed $SpoVT_C$ (red), $CodY_N$ (blue), and C-terminal domain of *T. maritima* IclR (green). The Zn atom bound to *T. maritima* IclR is shown as a gray sphere, and isoleucine bound to $CodY_N$ is shown as gold sticks. (d) Ribbon presentation of the potential ligand binding pocket of $SpoVT_C$ with relevant residues shown as sticks in element color code and labels.

amino acids appearing to be attached to the binding pocket rather than fully enclosed.

SpoVT has been crystallized as the apo protein, and the ligand binding pocket is in a closed conformation as the long loop L_S entirely covers the putative binding site (Figs. 2a and 4a). Various attempts at identifying ligands *in vitro* by testing cGMP, GDP, GTP, cAMP, dipicolinic acid, isoleucine, leucine, and valine as putative ligands have so far been unsuccessful. Due to the small forespore volume that may imply a high ligand concentration and thus probably a lower affinity, identification *in*

vitro is challenging, and it is tempting to speculate that an unusual or even a yet unidentified ligand might be the activation partner. Since the binding pocket is not conserved throughout SpoVT orthologs (Fig. 1), conserved patches occur in the α-helices (α2, α3, α4, and α5), in the loops connecting β-sheets β5 and β6 and between β8 and β9, and in β-sheets β8 and β9; hence, the SpoVT activation mode is probably not conserved in different organisms as well.

The GAF domain is a second dimerization module in SpoVT

The GAF domain forms the basis as the second oligomerization module through the formation of four-helix bundle interactions between helices α2 and α5, burying a surface of 610 Å^2.[19] The interface is formed as an anti-parallel four-helical coiled coil (Fig. 5a). The helices have pentadecad periodicity (i.e., 15 residue repeats over four helical turns, cor-

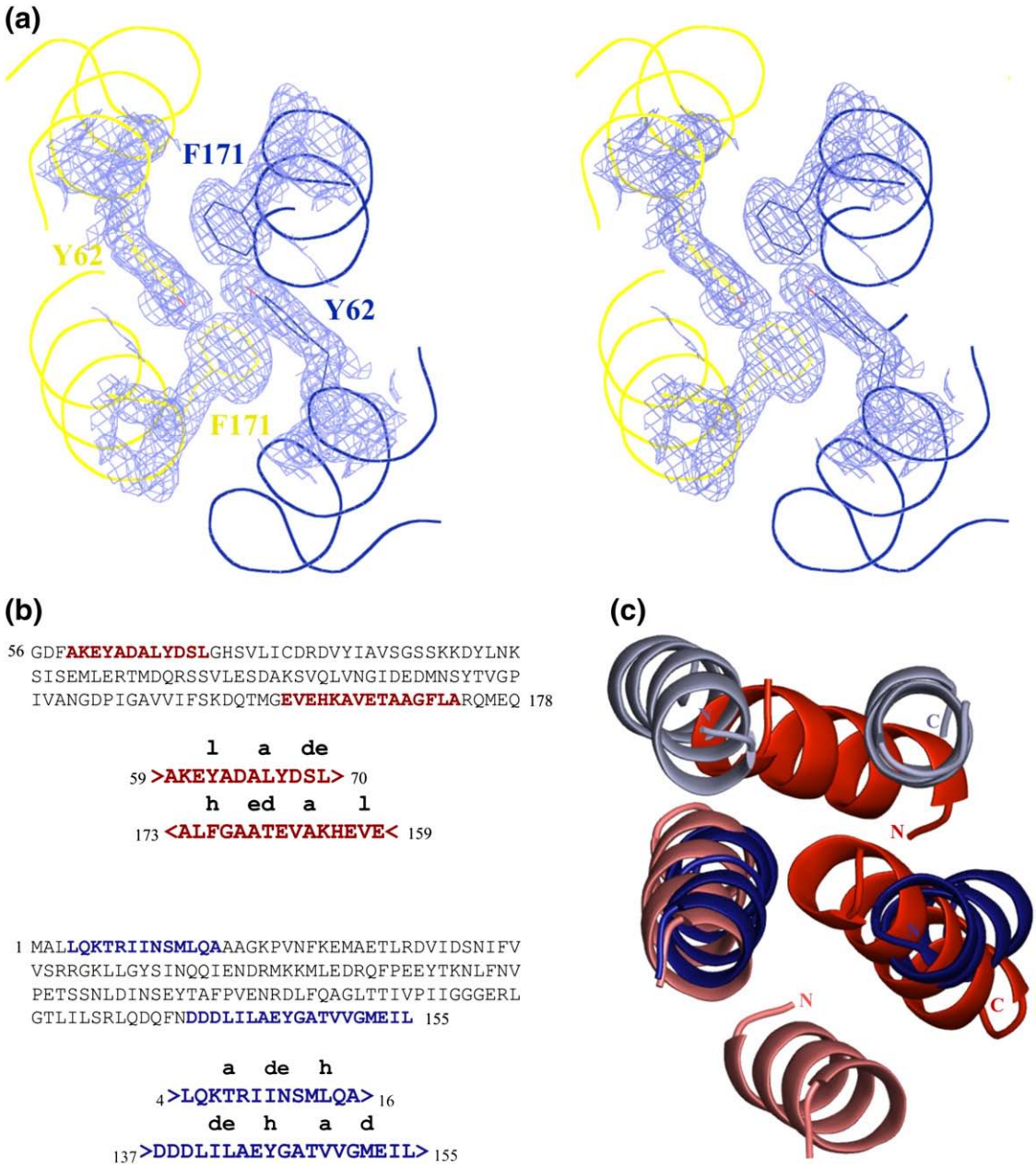

Fig. 5. Interaction of SpoVT$_C$ GAF domains. (a) Electron density of the interaction interface of SpoVT$_C$ domains contouring interacting Phe171 and Tyr62 residues in a stereo presentation. Residues are indicated and monomers are color coded. (b) Amino acid sequence of SpoVT$_C$ with residues shaping helices of the interaction interface highlighted in bold red (top). Amino acids of the helices that adopt hydrophobic core positions in the four-helical coiled-coil dimer interface are labeled (*a*, *d*, *e*, *h*, and *l*; bottom). The amino acid sequence of CodY$_N$ with residues shaping helices of the dimer interface is highlighted in bold blue (top). Residues in hydrophobic core positions are indicated (*a*, *d*, *e*, and *h*; bottom). (c) Superposition of the four-helical coiled-coil dimer interfaces of SpoVT$_C$ (red) and CodY$_N$ (blue). Monomers are depicted in different color intensities, and N- and C-termini are indicated.

responding to 3.75 residues per turn; repeat positions are denoted *a*–*o*, with hydrophobic core positions at *a*, *d*, *e*, *h*, and *l*). This causes a right-handed supercoil with helix crossing angles between 25° and 45°. The cores of the helices are not in register, but offset, as shown in Fig. 5b. This results in unusual core geometries, including two layers with complementary x-da packing:[27] Leu66 (*a*)–Ala164 (*a*)/Thr167 (*d*) and Ser69 (*d*)/Leu70 (*e*)–Ala168 (*e*). The bundle is not uniform in cross-section, being closer to a square at one end and to a rhombus at the other. The vectors connecting the central axes of symmetry-related helices have a length ratio of 16 to 15 Å in the first core layer (Tyr62–Phe171) and that of 20 to 10 Å in the last (Ser69/Leu70–Ala164). Interactions between Tyr62 and Phe171 are demonstrated in stereo in Fig. 5a.

As in other proteins,[12,13,17] the SpoVT GAF domain contributes to multimerization, but its geometry of interaction has not been observed elsewhere so far. When compared with its closest relative of known structure, $CodY_N$, many differences in helical packing geometry at the interface become apparent (Figs. 4b and 5c). $CodY_N$ also dimerizes *via* helical interactions, but, here, the dimer interface is an anti-parallel six-helix bundle with a buried surface of 1050 Å^2.[16] Each monomer contributes three helices, forming an anti-parallel, three-helical coiled coil. Two helices from each monomer build the dimer interface as a parallel four-helical coiled coil (Fig. 5c). All helices have hendecad periodicity (i.e., 11 residue repeats over three helical turns, corresponding to 3.67 residues per turn). This causes a minimal right-handed supercoil with helix crossing angles close to 0°. The cores of the helices are not in register, but offset, as shown in Fig. 5b; this is true for both the intramolecular three-helical coiled coils (data not shown) and the intermolecular four-helical coiled coil. The coiled coils are fairly uniform in cross-section; the vectors connecting the axes of symmetry-related helices in the central intermolecular coiled coil have a length ratio of about 14 to 17 Å.

The geometry of interaction between the SpoVT GAF domains clearly plays an important role in the overall shape of SpoVT and thus presumably in the functional properties of the protein as well.

SpoVT forms an overall asymmetric tetramer

The oligomer formation of SpoVT has already been proposed based on biochemical data describing possibly forms of dimers to tetramers.[22] By contrast, in our gel-filtration experiments, SpoVT eluted as a single peak with an apparent molecular mass of ~80 kDa, indicating a tetrameric protein complex (Fig. 6a). The monomeric molecular mass of 20.6 kDa for SpoVT in the complex was confirmed by SDS-PAGE analysis of the peak fraction (Fig. 6b).

Intermolecular electrostatic interactions between Arg40 and Glu54 and/or Glu35 and residues related by 2-fold symmetry (Fig. 6d) keep the SpoVT tetramer in the asymmetric conformation. The two domains in $SpoVT_A$ approximate each other such that the arginines Glu35 and Glu54 can interact with Arg40 (Fig. 6d). The $SpoVT_N$ domains in monomers A and B are shifted by about 90° relative to each other (Figs. 2e and 6c) and, together with intramolecular salt bridges (Fig. 2c), are responsible for the overall twist of the tetrameric complex (Fig. 6c). The residues involved in these electrostatic interactions are highly conserved among SpoVT orthologs (Fig. 1).

Overall, this results in an impressive asymmetry of the structure that is visible in both the AU and furthermore in the tetrameric complex structure using crystallographic symmetry (Figs. 2b and e and 6c). The tetrameric ring-shaped complex is assembled by four protein–protein interfaces. Two monomers associate *via* their AbrB-like domains to build swapped-hairpin β-barrels at the core of the structure with the GAF-like domains extending in two opposite directions (Fig. 2b). The dimers further assemble through helical interactions between their distant GAF domains (Fig. 5a). Conserved residues mapped onto the GAF surface result in patches of residues involved in essentially these protein–protein interactions (Fig. 2d). An interface area of ~4560 Å^2 obtained by twice the sum of the interaction area between the $SpoVT_N$ domain (1670 Å^2) and the $SpoVT_C$ domain (610 Å^2) forms a stable SpoVT tetramer. This complex of four monomers (A, B, C, and D; Fig. 6c) shows a slightly wounded spiral architecture that results in a convex shape. The complex is characterized by an overall asymmetry due to an anti-parallel assembly of dimers, composed of two conformationally differing monomers (Fig. 6c). This asymmetry could originate from the crystallographic packing but may also preexist in solution. In the presence of a yet unknown ligand, this asymmetry might be impaired, possibly induced by a conformational change in the loop L_S shielding the binding pocket followed by an overall relaxation of this energetically unfavorable state.

The DNA-binding specificity of SpoVT may be described as "limited promiscuity," similar to AbrB;[28] in both cases, the target promoters lack clear consensus binding motifs.[7] For the best studied swapped-hairpin transcription factor AbrB, it is assumed that a specific topology of compatible DNA segments is important for DNA-binding specificity.[29] The ability of a certain DNA target sequence to alter its conformation is therefore critical in its ability to bind AbrB. This flexibility represents an important characteristic of DNA that has been determined to be a contributing factor in the ability of AbrB to bind its targets. While we assume that the nature of the DNA promoter bound to AbrB and possibly SpoVT is palindromic, we also point to the biophysical nature of such promoters with an increased flexibility in the DNA structure. We note, however, that no structure has been determined as yet for a swapped-hairpin β-barrel in complex with its cognate DNA and therefore that details of this interaction remain conjectural.

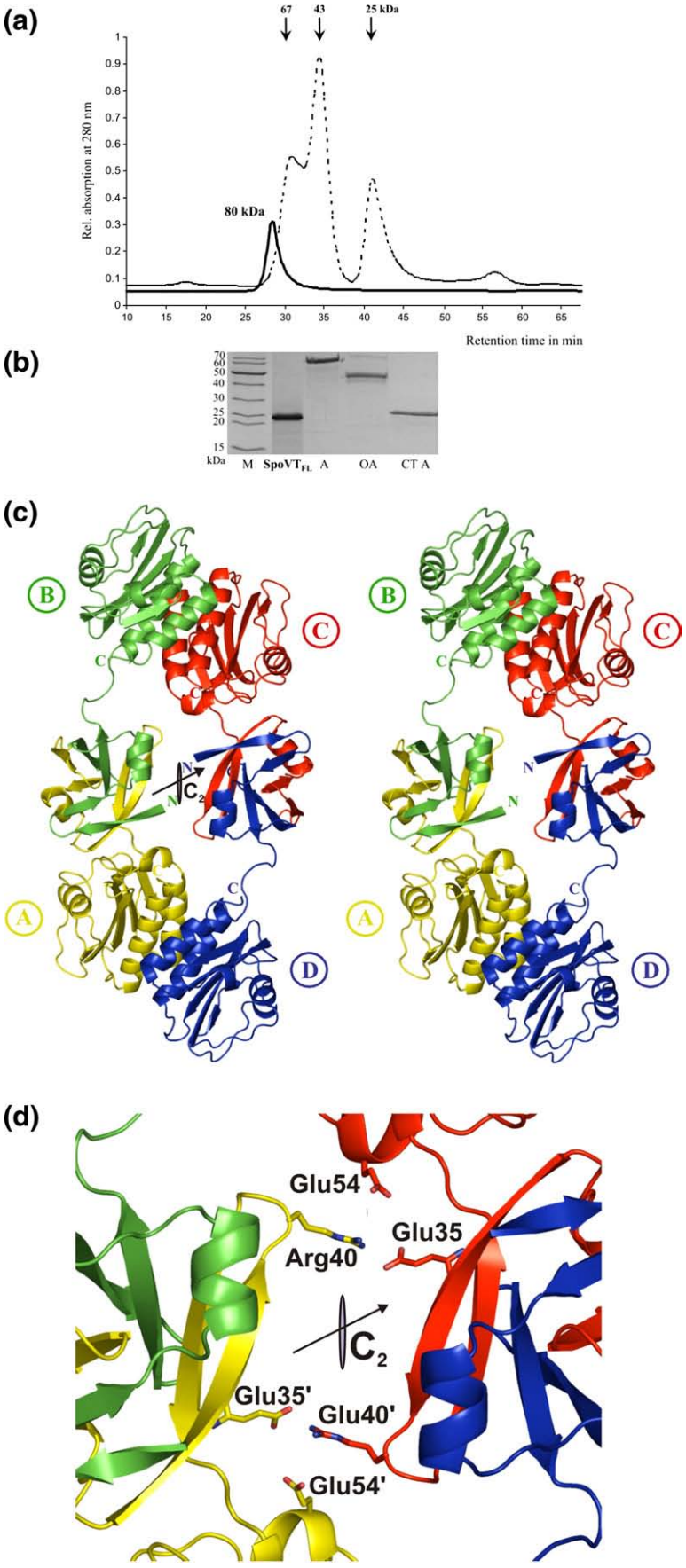

Fig. 6. Size-exclusion chromatography and overall fold of SpoVT. (a) Size-exclusion chromatography profile of SpoVT using a Superdex 75 column. Column calibration was done with albumin (A, 67 kDa), ovalbumin (OA, 43 kDa), and chymotrypsinogen A (CT A, 25 kDa). Elution profiles are illustrated as a black line for SpoVT and as a dotted line for the calibration protein mixture. (b) SDS-PAGE analysis of peak fractions is illustrated by the gel below with a specified molecular weight marker (M). (c) Stereo representation of the overall structure of tetrameric SpoVT colored by monomers (A, yellow; B, green; C, red; and D, blue). N- and C-termini together with the 2-fold axis are indicated. (d) Intermolecular electrostatic interactions between SpoVT$_A$ monomers with amino acids shown as sticks and labels. Monomers are colored similar to panel c.

Several dimeric transcription factors bind palindromic DNA sequences spanning a range of ~20 bp.[13,30,31] However, the two DNA-binding domains dock into two consecutive major grooves at a distance of ~34 Å in B-DNA. Swapped-hairpin β-barrels require tetramerization for such an interaction, since each DNA-binding domain is itself dimeric, and, indeed, AbrB-like transcription factors have been found to require tetramerization for their biological activity.[20,32] In the SpoVT crystal structure, the asymmetric SpoVT tetramer exposes two DNA-binding domains at a distance of ~37 Å, albeit with the putative DNA-binding clefts twisted against each other. Since SpoVT does not bind a subset of DNA segments (promoter regions of *bofC* and *spoVT*) *in vitro*, which agrees with experiments performed by others,[7,33] we assume that this conformation of the SpoVT tetramer represents an inactive form of the transcription factor. A rearrangement of the $SpoVT_N$ domains involving rotation and a reduction of their distance by about 3 Å appear necessary for binding to two consecutive major grooves of B-DNA. The rearrangement to a probably symmetric tetramer requires the disruption of the specific interactions between and within $SpoVT_A$ monomers. This transition is presumably triggered by ligands binding to the GAF domains inducing conformational changes as shown for $CodY_N$.[16] The role of the residues mediating these interactions will be tested by mutagenesis in the near future.

Conclusions

SpoVT is the first structurally characterized transcription factor that combines a swapped-hairpin β-barrel with a GAF domain. With the crystal structure of SpoVT, we show that the SpoVT GAF domain is involved in multimerization, resulting in a tetrameric state known to be essential for the DNA-binding activity of swapped-hairpin transcription factors. However, DNA binding by SpoVT *in vitro* can only be detected, albeit with loss of specificity, when $SpoVT_C$ is deleted,[7,33] indicating a regulatory role of $SpoVT_C$, presumably in response to a signaling molecule. Since ligand binding to a GAF domain causes large conformational changes throughout the domain,[16] we suppose ligand binding to the SpoVT GAF domains would disrupt the contacts stabilizing the asymmetric form of the SpoVT tetramer, allowing for a rearrangement in the position of the two DNA-binding domains and an interaction with two consecutive major grooves of B-DNA.

In summary, we propose that the asymmetric and unliganded tetramer described here represents the inactive state of SpoVT and that an active symmetric state can be induced only by a ligand specifically binding to the GAF domains. This ligand is as yet unknown and presumably occurs in increasing concentrations during forespore maturation, eventually triggering SpoVT activation upon crossing a concentration threshold.

Materials and Methods

Sequence and structure analysis

Sequence searches were performed with PSI-Blast[34] on the non-redundant protein sequence database at the National Center for Biotechnology Information‡ and with HHpred[35] on sequences from the PDB, filtered at 70% sequence identity (PDB70). Sequence alignments were made with ClustalW.[36] Structure alignments were obtained from DALI[37] and were investigated further using the PISA database[19] and the RMSD calculator.[21] Molecular images were generated with PYMOL§.

Protein cloning, expression, and purification

The *spoVT* open reading frame was amplified from chromosomal DNA of *B. subtilis* PY79 by PCR (polymerase chain reaction) using the primers SpF 5′-GGAATTCCATATGCATCATCATCATCATCATAAAGCAACCGGTATCGTACGTCGTATTGATGACTTAGG and SpR 5′-CCGCTCGAGTTACTGTTCCATTTGACGAGCCAAAATCCAGCTGCGTTTCAACGGC to obtain the full-length protein (SpoVT, residues 1–178). For the C-terminal domain ($SpoVT_C$, residues 56–178), the primers were SpcF 5′-GGAATTCCATATGCATCATCATCATCATCATGGAGACTTTGCAAAGGAGTATGCAGACGCGCTTTAGCAGAGCC and SpR, as described above. Each construct contained an N-terminal His_6 tag. The two gene products were cloned into pET30b(+) vectors (Novagen).

Overexpression of recombinant proteins, for subsequent purification, was performed in *E. coli* C41(DE3) strain (OverExpress). Cells were grown in LB medium containing 50 μg ml^{-1} of kanamycin, and expression was induced by addition of 1 mM IPTG (isopropyl-1-thio-β-D-*g*alactopyranoside). Selenomethionine (Se-Met)-labeled proteins were expressed in *E. coli* B834(DE3) cells (Novagen) grown in M9 minimal medium[38] containing 4 mg l^{-1} of Se-Met instead of Met.

Cell lysis was enforced using the French press method, and insoluble cell particles were removed by centrifugation. The His_6-tagged proteins were purified from *E. coli* lysate under native conditions using Ni–NTA Superflow (Qiagen) affinity matrix following the manufacturer's protocol‖. Eluted fractions were analyzed by SDS-PAGE,[39] and the ones containing the proteins of interest were pooled together. For crystallization experiments, both proteins were dialyzed against 20 mM Mops (3-morpholinopropane-1-sulfonic acid) buffer, pH 7.25, and 120 mM NaCl followed by preparative size-exclusion chromatography using a Sephadex G-75 column (GE Healthcare) to achieve sufficient purity. The protein solutions were adjusted to final concentrations of 10 and 20 mg ml^{-1} *via* ultrafiltration (Amicon) and applied to crystallization trials.

Size-exclusion chromatography

Analytical size-exclusion chromatography was performed on a SMART system using a Superdex 75 column

‡ http://www.ncbi.nlm.nih.gov
§ http://pymol.sourceforge.net/
‖ http://www1.qiagen.com/literature/handbooks/literature.aspx?id=1000146

PC 3.2/30 (Amersham Biosciences). The experiments were done in 20 mM Mops buffer, pH 7.25, and 120 mM NaCl at a flow rate of 40 μl min^{-1} with a protein concentration of 1.5 mg ml^{-1}. Eighty-microliter fractions were collected for subsequent SDS-PAGE analysis. Calibration was carried out with proteins of definite mass (LMW Calibration Kit, Amersham Biosciences).

Protein crystallization

The proteins were crystallized by vapor diffusion in hanging drops at 291 K containing a 1:1 volume ratio of either 10 or 20 mg ml^{-1} of protein in chromatography buffer and crystallization buffer. Crystallization conditions were screened with commercially available reagents (Hampton Research, Jena Bioscience, Emerald BioSystems, and Sigma). Crystals appeared after several days. The crystals of SpoVT were grown with the addition of 1 M NaCl, 100 mM cacodylate, pH 6.5, 30% (v/v) PEG (polyethylene glycol) 600, and 10% (v/v) glycerol (Emerald BioSystems, Cryo Screen II, condition 6). $SpoVT_C$ formed crystals in 200 mM NaCl, 100 mM Hepes [4-(2-*h*ydroxy*e*thyl)-1-*p*iperazine*e*thane*s*ulfonic acid], pH 7.5, and 25% (w/v) PEG 3350 (Hampton Research, Index Screen, condition 72), while the Se-Met derivate crystals appeared in 200 mM $MgCl_2$, 100 mM Hepes, pH 7.5, and 25% (w/v) PEG 3350 (Hampton Research, Index Screen, condition 84). Optimization of preliminary crystallization conditions was achieved by performing additive screens (Hampton Research, Additive Screen). In the case of Se-Met $SpoVT_C$, the presence of 3% (w/v) dextran sulfate led to the formation of crystals of a different morphology after an incubation period of 3.5 months.

Data collection

All diffraction data were collected at the Swiss Light Source beamline X10SA PXII at the Paul Scherrer Institut in Villigen, Switzerland. The X-ray data of the native and Se-Met $SpoVT_C$ crystals were acquired at a wavelength of 0.979 Å; those of native SpoVT crystals, at a wavelength of 0.978 Å. The data were recorded on a mar225 CCD detector. Single protein crystals were flash frozen in liquid nitrogen, and image data were collected at 100 K. The $SpoVT_C$ crystals were protected by the addition of 5% (v/v) PEG 200, and 280 1-deg images were collected. For the Se-Met derivate, a SAD experiment at the selenium absorption edge was performed. The native SpoVT crystals were flash frozen directly, and data of 280° were collected.

For SpoVT, a resolution of 2.6 Å was obtained, while the protein crystals of $SpoVT_C$ yielded a resolution of 1.5 Å. The initial crystals of the $SpoVT_C$ Se-Met derivate diffracted poorly (to only 3.5 Å) and did not allow for solution of the structure. $SpoVT_C$ Se-Met crystals of a different morphology, gained by optimizing crystallization conditions, diffracted to a resolution of 2.1 Å, which enabled tracing of the structure. For complete crystallographic analysis results, see Table 1.

Data processing, structure determination, and refinement

The SAD data of the Se-Met $SpoVT_C$ crystal were integrated, merged, and scaled using the program package XDS/XSCALE.[40] The intensity data were converted into amplitudes using TRUNCATE.[41] Six Se sites were located and experimental phases were calculated using SOLVE/RESOLVE.[42,43] A partial model was built automatically with RESOLVE;[44] missing parts of the protein were built manually and later refined against the native data set. The initial model was further refined by several cycles of model rebuilding with COOT[45] and automatic refinement using REFMAC.[46] Final refinement was performed with CNS[47] and PHENIX.[48] Ninety-eight solvent water molecules were added using the ARP/wARP program.[49] The N-terminal His_6 tag and the last five residues of the protein domain were absent in the $SpoVT_C$ model. The refined $SpoVT_C$ model was used as a search model for the structure determination of SpoVT by molecular replacement. The data of the native SpoVT protein crystal were processed as described above and the structure was solved using MOLREP.[50] To improve the density of the N-terminal domain, which initially could not be placed by molecular replacement using the NMR structure of AbrB,[9] we applied the prime-and-switch routine of the RESOLVE program. The SpoVT structure was completed by density fitting the N-terminal part of AbrB (PDB entry 1YFB) into the solvent-flattened and averaged map. Residues different between AbrB and SpoVT were exchanged and the model was refined with non-crystallographic and twin-lattice symmetry restraints by iterative cycles of model building and refinement with COOT, REFMAC, and PHENIX. The geometry was finally checked with PROCHECK.[51] The model consists of four chains with residues 3–177 for chains A and C and residues 5–177 for chains B and D. The X-ray data statistics are listed in Table 1.

PDB depositions

The atomic coordinates for the crystal structures of SpoVT and $SpoVT_C$ have been deposited in the PDB under accession numbers 2W1R and 2W1T, respectively.

Acknowledgements

This work was supported by institutional funds from the Max Planck Society and a Deutsche Forschungsgemeinschaft grant given to K.Z. (ZE522/3-3). We thank the staff at the Swiss Light Source (Paul Scherrer Institut, Villigen, Switzerland) and Dr. D. Oesterhelt of the Max Planck Institute for Biochemistry for continuous interest. Special thanks go to P. Reichelt, Dr. M. Aivaliotis, and K. Konstantinidis for proofreading of the manuscript and to P. Wollmann, Dr. M. Grininger, and Dr. P. Johansson for advice and discussions.

References

1. Errington, J. (2003). Regulation of endospore formation in *Bacillus subtilis*. *Nat. Rev. Microbiol.* **1**, 117–126.
2. Wang, S. T., Setlow, B., Conlon, E. M., Lyon, J. L., Imamura, D., Sato, T. *et al.* (2006). The forespore line of gene expression in *Bacillus subtilis*. *J. Mol. Biol.* **358**, 16–37.
3. Steil, L., Serrano, M., Henriques, A. O. & Volker, U. (2005). Genome-wide analysis of temporally regulated

and compartment-specific gene expression in sporulating cells of *Bacillus subtilis*. *Microbiology*, **151**, 399–420.
4. Igarashi, T. & Setlow, P. (2006). Transcription of the *Bacillus subtilis* gerK operon, which encodes a spore germinant receptor, and comparison with that of operons encoding other germinant receptors. *J. Bacteriol.* **188**, 4131–4136.
5. Bagyan, I., Hobbot, J. & Cutting, S. (1996). A compartmentalized regulator of developmental gene expression in *Bacillus subtilis*. *J. Bacteriol.* **178**, 4500–4507.
6. Le Breton, Y., Mohapatra, N. P. & Haldenwang, W. G. (2006). *In vivo* random mutagenesis of *Bacillus subtilis* by use of TnYLB-1, a mariner-based transposon. *Appl. Environ. Microbiol.* **72**, 327–333.
7. Dong, T. C., Cutting, S. M. & Lewis, R. J. (2004). DNA-binding studies on the *Bacillus subtilis* transcriptional regulator and AbrB homologue, SpoVT. *FEMS Microbiol. Lett.* **233**, 247–256.
8. Strauch, M. A., Ballar, P., Rowshan, A. J. & Zoller, K. L. (2005). The DNA-binding specificity of the *Bacillus anthracis* AbrB protein. *Microbiology*, **151**, 1751–1759.
9. Coles, M., Djuranovic, S., Soding, J., Frickey, T., Koretke, K., Truffault, V. *et al.* (2005). AbrB-like transcription factors assume a swapped hairpin fold that is evolutionarily related to double-psi beta barrels. *Structure*, **13**, 919–928.
10. Söding, J., Biegert, A. & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, 244–248.
11. Aravind, L. & Ponting, P. (1997). The GAF domain: an evolutionary link between diverse phototransducing proteins. *Trends Biochem. Sci.* **22**, 458–459.
12. Martinez, S. E., Wu, Y. A., Glavas, N. A., Tang, X. B., Turley, S., Hol, W. G. J. & Beavo, J. A. (2002). The two GAF domains in phosphodiesterase 2A have distinct roles in dimerization and in cGMP binding. *Proc. Natl Acad. Sci. USA*, **99**, 13260–13265.
13. Vannini, A., Volpari, C., Gargioli, C., Muraglia, E., Cortese, R., De Francesco, R. *et al.* (2002). The crystal structure of the quorum sensing protein TraR bound to its autoinducer and target DNA. *EMBO J.* **21**, 4393–4401.
14. Zhang, K. Y., Card, G. L., Suzuki, Y., Artis, D. R., Fong, D., Gillette, S. *et al.* (2004). A glutamine switch mechanism for nucleotide selectivity by phosphodiesterases. *Mol. Cell*, **15**, 279–286.
15. Handke, L. D., Shivers, R. P. & Sonenshein, A. L. (2008). Interaction of *Bacillus subtilis* CodY with GTP. *J. Bacteriol.* **190**, 798–806.
16. Levdikov, V. M., Blagova, E., Joseph, P., Sonenshein, A. L. & Wilkinson, A. J. (2006). The structure of CodY, a GTP- and isoleucine-responsive regulator of stationary phase and virulence in Gram-positive bacteria. *J. Biol. Chem.* **281**, 11366–11373.
17. Ho, Y. S., Burden, L. M. & Hurley, J. H. (2000). Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. *EMBO J.* **19**, 5288–5299.
18. Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.
19. Krissinel, E. & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797.
20. Bobay, B. G., Andreeva, A., Mueller, G. A., Cavanagh, J. & Murzin, A. G. (2005). Revised structure of the AbrB N-terminal domain unifies a diverse superfamily of putative DNA-binding proteins. *FEBS Lett.* **579**, 5669–5674.
21. Claude, J. B., Suhre, K., Notredame, C., Claverie, J. M. & Abergel, C. (2004). CaspR: a Web server for automated molecular replacement using homology modelling. *Nucleic Acids Res.* **32**, 606–609.
22. Vaughn, J. L., Feher, V., Naylor, S., Strauch, M. A. & Cavanagh, J. (2000). Novel DNA binding domain and genetic regulation model of *Bacillus subtilis* transition state regulator abrB. *Nat. Struct. Biol.* **7**, 1139–1146.
23. Loris, R., Marianovsky, I., Lah, J., Laeremans, T., Engelberg-Kulka, H., Glaser, G. *et al.* (2003). Crystal structure of the intrinsically flexible addiction antidote MazE. *J. Biol. Chem.* **278**, 28252–28257.
24. Kamada, K., Hanaoka, F. & Burley, S. K. (2003). Crystal structure of the MazE/MazF complex: molecular bases of antidote–toxin recognition. *Mol. Cell*, **11**, 875–884.
25. Martinez, S. E., Bruder, S., Schultz, A., Zheng, N., Schultz, J. E., Beavo, J. A. & Linder, J. U. (2005). Crystal structure of the tandem GAF domains from a cyanobacterial adenylyl cyclase: modes of ligand binding and dimerization. *Proc. Natl Acad. Sci. USA*, **102**, 3082–3087.
26. Zhang, R. G., Kim, Y., Skarina, T., Beasley, S., Laskowski, R., Arrowsmith, C. *et al.* (2002). Crystal structure of *Thermotoga maritima* 0065, a member of the IclR transcriptional factor family. *J. Biol. Chem.* **277**, 19183–19190.
27. Hulko, M., Berndt, F., Gruber, M., Linder, J. U., Truffault, V., Schultz, A. *et al.* (2006). The HAMP domain structure implies helix rotation in transmembrane signaling. *Cell*, **126**, 929–940.
28. Bobay, B. G., Mueller, G. A., Thompson, R. J., Murzin, A. G., Venters, R. A., Strauch, M. A. & Cavanagh, J. (2006). NMR structure of AbhN and comparison with AbrBN: first insights into the DNA binding promiscuity and specificity of AbrB-like transition state regulator proteins. *J. Biol. Chem.* **281**, 21399–21409.
29. Bobay, B. G., Benson, L., Naylor, S., Feeney, B., Clark, A. C., Goshe, M. B. *et al.* (2004). Evaluation of the DNA binding tendencies of the transition state regulator AbrB. *Biochemistry*, **43**, 16106–16118.
30. Jordan, S. R. & Pabo, C. O. (1988). Structure of the lambda complex at 2.5 Å resolution: details of the repressor–operator interactions. *Science*, **242**, 893–899.
31. Baleja, J. D., Anderson, F. & Sykes, B. D. (1991). Different interactions of Cro repressor dimer with the left and right halves of OR3 operator DNA. *J. Biol. Chem.* **266**, 22115–22124.
32. Yao, F. & Strauch, M. A. (2005). Independent and interchangeable multimerization domains of the AbrB, Abh, and SpoVT global regulatory proteins. *J. Bacteriol.* **187**, 6354–6362; [erratum appears in J. Bacteriol. 2005 (Nov), 187, 7546].
33. Benson, L. M., Vaughn, J. L., Strauch, M. A., Bobay, B. G., Thompson, R., Naylor, S. & Cavanagh, J. (2002). Macromolecular assembly of the transition state regulator AbrB in its unbound and complexed states probed by microelectrospray ionization mass spectrometry. *Anal. Biochem.* **306**, 222–227.
34. Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
35. Söding, J. (2005). Protein homology detection by HMM–HMM comparison. *Bioinformatics*, **21**, 951–960.
36. Higgins, D., Thompson, J., Gibson, T., Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W:

improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.
37. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.
38. Doublie, S. (2007). Production of selenomethionyl proteins in prokaryotic and eukaryotic expression systems. *Methods Mol. Biol.* **363**, 91–108.
39. Hames, B. D. & Rickwood, D. (1990). *Gel Electrophoresis of Proteins: A Practical Approach*, 2nd edit. Oxford University Press, New York, NY.
40. Kabsch, W. (1993). *J. Appl. Crystallogr.* **26**, 795–800.
41. French, G. S. & Wilson, K. S. (1978). *Acta Crystallogr. Sect. A*, **34**, 517.
42. Terwilliger, T. C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **55**, 849–861.
43. Terwilliger, T. C. (2000). Maximum likelihood density modification. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **56**, 965–972.
44. Terwilliger, T. C. (2003). Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **59**, 38–44.
45. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
46. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **53**, *240–255.*
47. Brünger, A. T., Adams, P. D. & Rice, L. M. (1998). Recent developments for the efficient crystallographic refinement of macromolecular structures. *Curr. Opin. Struct. Biol.* **5**, 606–611.
48. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W. *et al.* (2002). PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **58**, 1948–1954.
49. Perrakis, A., Harkiolaki, M., Wilson, K. S. & Lamzin, V. S. (2001). ARP/wARP and molecular replacement. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **57**, 1445–1450.
50. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025.
51. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.