Structure, Vol. 13, 919–928, June, 2005, ©2005 Elsevier Ltd All rights reserved. DOI 10.1016/j.str.2005.03.017

AbrB-like Transcription Factors Assume a Swapped Hairpin Fold that Is Evolutionarily Related to Double-Psi β Barrels

Murray Coles,[1,3] Sergej Djuranovic,[1,3] Johannes Söding,[1] Tancred Frickey,[1] Kristin Koretke,[2] Vincent Truffault,[1] Jörg Martin,[1] and Andrei N. Lupas[1,*]
[1]Department of Protein Evolution
Max-Planck-Institute for Developmental Biology
72076 Tübingen
Germany
[2]Protein Bioinformatics Group
GlaxoSmithKline
Collegeville, Pennsylvania 19426

Summary

AbrB is a key transition-state regulator of *Bacillus subtilis*. Based on the conservation of a βαβ structural unit, we proposed a β barrel fold for its DNA binding domain, similar to, but topologically distinct from, double-psi β barrels. However, the NMR structure revealed a novel fold, the "looped-hinge helix." To understand this discrepancy, we undertook a bioinformatics study of AbrB and its homologs; these form a large superfamily, which includes SpoVT, PrlF, MraZ, addiction module antidotes (PemI, MazE), plasmid maintenance proteins (VagC, VapB), and archaeal PhoU homologs. MazE and MraZ form swapped-hairpin β barrels. We therefore reexamined the fold of AbrB by NMR spectroscopy and found that it also forms a swapped-hairpin barrel. The conservation of the core βαβ element supports a common evolutionary origin for swapped-hairpin and double-psi barrels, which we group into a higher-order class, the cradle-loop barrels, based on the peculiar shape of their ligand binding site.

Introduction

In switching from exponential growth to stationary phase, *Bacillus subtilis* may induce a number of functions designed to insure survival in a more hostile environment. Among them are biofilm formation, antibiotic production, motility, development of competence for DNA uptake, synthesis of extracellular enzymes, and sporulation (Strauch and Hoch, 1993; Phillips and Strauch, 2002). The phase in which the cell decides on its response to a degrading environment is called the transition state; it is dominated by the activity of a heterogeneous class of transcription factors, called transition-state regulators. One such key regulator in *B. subtilis* is AbrB (antibiotic resistance protein B), a protein whose homologs appear to be broadly present in bacterial species (Strauch et al., 1989; Klein and Marahiel, 2002; Shafikhani and Leighton, 2004; Hamon et al., 2004). AbrB binds to DNA via its N-terminal 50 or so residues (AbrB-N) (Xu and Strauch, 2001). Despite being known to regulate the expression of at least 60 different genes, its promotor recognition consensus sequence and mode of interaction with DNA remain unclear (Strauch, 1995; Vaughn et al., 2000, 2001; Cavanagh et al., 2002; Benson et al., 2002; Bobay et al., 2004).

During a study of the substrate-recognition domain of the AAA ATPase VAT, we found evidence for an evolutionary connection between double-psi β barrels and AbrB-N (Coles et al., 1999). VAT is the Cdc48 homolog from the archaeon *Thermoplasma acidophilum* (Pamnani et al., 1997), whose N-terminal, substrate-recognition domain (VatN) is formed of two subdomains, VatN-N and VatN-C. The former is a double-psi barrel, a complicated, six-stranded structure consisting of two homologous halves, whose name derives from the fact that strands β1 and β2 of each half are connected by a loop that passes over the symmetry-related strand β2′, resembling the Greek letter Ψ in top view (Castillo et al., 1999). In an attempt to understand the evolutionary origin of such a topologically complex fold, we searched the sequence databases for possible precursor forms with a simpler topology, and we identified AbrB as a distant homolog in the process. The similarity between the two proteins hinged on the presence of a conserved sequence element, which in VatN-N formed two β strands flanking an α helix and enclosing an orthogonal turn with a conspicuous Gly-Asp motif (referred to in the following as the GD box). This element appeared to be elaborated by an additional N-terminal β strand in each of the two symmetry-related halves of VatN-N and by a C-terminal strand in AbrB-N, leading us to propose that the two proteins were related by circular permutation (for an explanation of circular permutation in proteins, see Grishin, 2001). Correspondingly, we predicted AbrB-N to form a β barrel that resembled double-psi barrels but had a distinct topology lacking the two psi loops (Coles et al., 1999), thus corresponding to the simpler precursor fold we had been searching for.

In 2000, Cavanagh and coworkers published the solution structure of AbrB-N (Vaughn et al., 2000), which differed substantially from the one predicted by comparison to VatN-N. Instead of a barrel, AbrB-N appeared as a side-by-side dimer of two three-stranded β meanders, with two short helices placed equatorially and connected to the meanders via long loops. The authors called this novel fold the "looped-hinge helix." This fold was incompatible with the proposed evolutionary scenario, since the structure of the core βαβ element was not preserved. In order to understand this discrepancy, we undertook a theoretical and experimental reevaluation of AbrB-N. We found that AbrB is a representative of a large superfamily of known and putative transcription factors and that its fold is a β barrel consisting of two pairs of interleaved β hairpins, in

*Correspondence: andrei.lupas@tuebingen.mpg.de
[3]These authors contributed equally to this work.

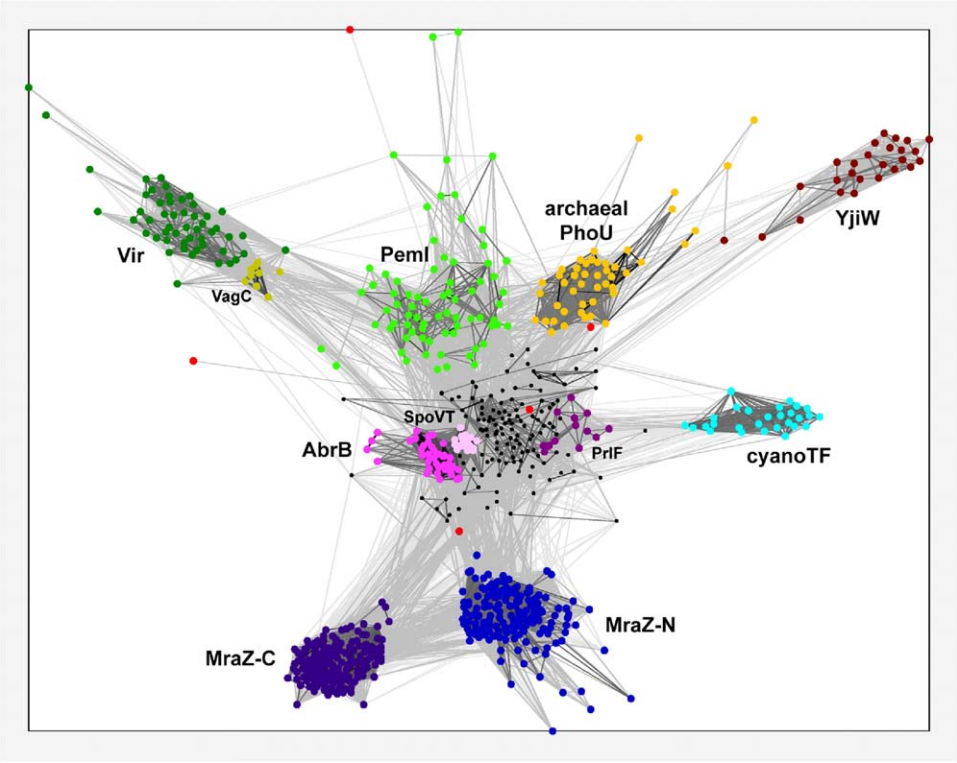

Figure 1. Cluster Analysis of the AbrB Superfamily

The eight main clusters are highlighted in different colors and explained in the text. The central AbrB core cluster is uncolored, except for the three subclusters corresponding to AbrB, SpoVT, and PrlF, which are labeled. In the Vir cluster, the VagC subcluster is highlighted in a different color and labeled. Five sequences that were not assigned to any cluster are colored red and are explained in the text. The role of incorrectly assigned start codons in generating apparent outliers (e.g., in the MraZ-N, but not the MraZ-C, cluster) is discussed in the Experimental Procedures.

agreement with the crystal structures of two homologs and in contradiction to the published AbrB structure.

Results

A Superfamily of AbrB-like Transcription Factors

In a first step, we searched the current protein sequence database for homologs of AbrB, by using a new software tool, HHsenser, which we obtained by combining our method for the comparison of profile Hidden Markov Models (HHpred) (Soding, 2004) with a sensitive search routine based on PSI-Blast (SENSER; see Experimental Procedures) (Koretke et al., 2002). We identified 724 sequences resembling AbrB-N in 580 unique proteins. A cluster analysis of these sequences by using CLANS (Frickey and Lupas, 2004) revealed the existence of eight major groups (Figure 1), and sequence comparisons confirmed their similarity to AbrB-N (Figure 2). The groups are as follows.

AbrB Core Group

This group contains two well-defined subclusters, one formed by AbrB itself and its closest relatives, and the other by SpoVT homologs. SpoVT is a late sporulation factor that modulates forespore-specific, σ^G-dependent transcription (Bagyan et al., 1996). AbrB and SpoVT share 65% sequence identity in their N domains, but they seem to have clearly distinct DNA binding activities, since AbrB-N cannot substitute for the homologous domain in SpoVT (Dong et al., 2004). In contrast to AbrB, which has a short, largely unstructured C-terminal region, SpoVT has a folded C-terminal domain of about 125 residues, which is essential for the structure and function of the N domain (Dong et al., 2004). Using HHpred, we find that SpoVT-C is most likely a GAF or PAS domain and thus the effector site for an as yet unidentified small molecule in the progression of sporulation.

AbrB and SpoVT are embedded into a large and diffuse group of sequences from bacteria and archaea, which also include PrlF homologs of proteobacteria. PrlF, an enhancer of Lon protease activity (Snyder and Silhavy, 1992), is also known as suppressor of HtrA protease (SohA), due to its ability to suppress the phenotype of temperature-sensitive HtrA (DegP) mutants and thus influence the stress response (Baird and Georgopoulos, 1990) (its name was changed to "HtaR suppressor" in the annotation of Synechocystis PCC6803, but we are not aware of an HtaR gene). In most genomes, PrlF is translationally coupled to a homolog of *E. coli* YhaV, which shows distant but significant similarity to RelE according to HHpred. RelE is part of a toxin-antitoxin system (together with the ribbon-helix-helix transcription factor RelB) (Gotfredsen and Gerdes, 1998); it is a global inhibitor of translation, which

Figure 2. Multiple Alignment of AbrB-like Sequences

The sequences are grouped by cluster and are listed in the order in which they appear in the text. The consensus secondary structure for each group is listed above the sequences (uppercase, observed; lowercase, predicted by Psipred; H/h, helix; S/s, strand). Residues forming the hydrophobic core of the barrel are colored blue; the signature residues of the PxxxR motif and the GD box are colored red, as are the positive charges in the β1–β2 hairpin, which are likely to contribute to DNA binding. The sequences are: AbrB core group: *B. subtilis* AbrB (gi113009), *B. subtilis* SpoVT (gi586883), *E. coli* PrlF (gi134687), *Archaeoglobus fulgidus* AF2359 (gi11499936), *Sulfolobus solfataricus* SSO5984 (gi18272484); Vir group: *Dichelobacter nodosus* VapB (gi7388354), *Shigella flexneri* MvpT (gi32307073), *Actinobacillus actinomycetemcomitans* VppA (gi3786344), *Salmonella dublin* VagC (gi7388352); PemI group: *E. coli* PemI (gi41057026), *E. coli* MazE (gi126777), *E. coli* ChpBI (gi2144940), *Neisseria meningitidis* NMB0914 (gi15676809); archaeal PhoU group: *A. fulgidus* AF0472 (gi11498083), *Pyrococcus furiosus* PF0141 (gi18976513), *Sulfolobus tokodaii* ST1484 (gi15921777), *Pyrobaculum aerophilum* PAE2220 (gi18313189); MraZ-N and MraZC groups: *Mycoplasma pneumoniae* MraZ (gi2496334), *E. coli* MraZ (gi140163); YjiW group: *E. coli* YjiW (gi732102), *Erwinia carotovora* ECA2121 (gi50121049) and ECA2120 (gi50121048); cyanobacterial ORFs: *Crocosphaera watsonii* Cwat03006552 (gi46118132), *Synechocystis* PCC6803 sll0822 (gi16331736), *Prochlorococcus marinus* Pro0575 (gi33240026); ungrouped: *Geobacter metallireducens* RecG (gi48846167), *Lactococcus lactis* YkiI (permease) (gi15673060), *Aeropyrum pernix* IlvH (gi14602188), *Pyrococcus abyssi* PAB0691 (IlvH) (gi14521238), *Thermoplasma acidophilum* VatN-N domain (gi11387127).

cleaves ribosome-associated transcripts and may be related to eukaryotic nonsense-mediated RNA decay (Anantharaman and Aravind, 2003). Correspondingly, PrlF and YhaV may represent a toxin-antitoxin system involved in the regulation of translation. Toxin-antitoxin operons appear to be involved in the maintenance of plasmids ("addiction modules"), as well as in the regulation of macromolecular synthesis during nutritional stress and in the programmed cell death of prokaryotes (Engelberg-Kulka and Glaser, 1999; Gerdes, 2000). The labile antitoxin, which also serves as the transcription regulator for the operon, inhibits the stable toxin through formation of a complex. Upon degradation of the antitoxin (for example, after loss of the plasmid in addiction systems), the toxin exerts its activity, frequently killing the cell.

Maintenance Proteins of Virulence Plasmids

This group contains several transcription factors (antitoxins) of plasmid maintenance systems (postsegregational killing systems) and of pathogenicity islands. Among them are VapB of *Dichelobacter nodosus* (Katz et al., 1992), MvpT of *Shigella flexneri* (Radnedge et al., 1997), VppA of *Actinobacillus actinomycetemcomitans* (Mayer et al., 1999), and VagC of *Salmonella* spp. (Pullinger and Lax, 1992) (which forms a distinct subcluster within this group). The toxins of these systems are all related to the PilT-N terminal (PIN) domain and consist of a flavodoxin-like, doubly wound parallel β sheet with strand order 2134 (PDB code 1O4W). They are most likely nucleases homologous to the 5′ to 3′ exonuclease domain of Taq polymerase and are grouped with it in superfamily c.120.1. in the Structural Classification of Proteins (SCOP) database (Andreeva et al., 2004) (http://scop.mrc-lmb.cam.ac.uk/scop/index.html).

Antidotes of Addiction Modules

This group is also formed of transcription factors from toxin-antitoxin systems, but the toxins belong neither to the RelE nor to the PIN class. Most of the addiction modules, such as PemIK, MazEF (also known as ChpRA or ChpAIK), kis/kid (ParD), and ChpBIK (all from

E. coli) act through toxins with an SH3-like barrel fold (Kamada et al., 2003). Of these, PemK and MazF have been shown to be endonucleases that cleave mRNAs at specific sequences (Zhang et al., 2003, 2004). A minority of the addiction modules in this group act through a different toxin, called death on curing (DOC) (Lehnherr et al., 1993). The structure of DOC domains is unknown, but they have also been predicted to have nuclease activity, based on their residue conservation patterns (Anantharaman and Aravind, 2003).

Identification of this group was of particular importance to our study, since the crystal structure of MazE has been determined, both alone (1MVF) (Loris et al., 2003) and in complex with MazF (1UB4) (Kamada et al., 2003). Its fold does not resemble that of AbrB in 1EKT, despite the clear homology of the two proteins. Rather, MazE forms a β barrel with many of the properties we had proposed by extrapolation from the double-psi barrel of VatN (Figure 3A).

Archaeal PhoU Homologs

This group is formed entirely of proteins from archaea, and its members share a conserved domain structure consisting of an N-terminal AbrB-like domain, a central domain of unknown activity, and two C-terminal PhoU elements, which come together into a single six-helical bundle with 2-fold pseudosymmetry, as seen in the crystal structures of PhoU from *Thermotoga maritima* (1SUM) and *Aquifex aeolicus* (1T72). The structure of the central domain is predicted to be IF3 like (SCOP: d.68) by HHpred. Like PhoU homologs in proteobacteria (Wanner, 1993), the archaeal proteins are usually found in phosphate-specific transport (pst) system operons. Since archaea lack the PhoRB two-component signal transduction system, which regulates the expression of the pst operon in bacteria, it seems possible that transcriptional regulation has been integrated into the archaeal PhoU protein via fusion to an AbrB-like domain.

The N- and C-Terminal Halves of MraZ Homologs

mraZ is the first gene in the division and cell wall (dcw) cluster, which comprises 16 genes in *E. coli* and whose gene order is highly conserved throughout bacteria (Vicente et al., 1998). Although most of the genes have been assigned a function, the role of *mraZ* remains unknown; our analysis suggests that it is a transcription factor involved in the expression of the dcw cluster. Structurally, MraZ takes a special place among AbrB homologs in that the two subunits needed to form the fold are fused here into a single chain. Since the N and C termini of the two subunits of AbrB are at opposite ends of the dimer, the distance must be bridged by a connecting sequence. The recently determined crystal structure of MraZ from *Mycoplasma pneumoniae* (1N0G) (Chen et al., 2004) shows that this is achieved via a helical hairpin (Figure 3A). The helical hairpin appears after each half of MraZ, and the two halves are nearest neighbors in the cluster map, suggesting that the protein originated through duplication and fusion of an ancestral homodimeric MraZ, which already contained the helical hairpin at its C terminus. The structure of MraZ, claimed to represent a novel fold (Chen et al., 2004), in fact closely resembles that of MazE.

Hypothetical Proteins of Proteobacteria

This group is formed almost exclusively of proteins from proteobacteria, with the exception of EF2302 of *Enterococcus faecalis*, and is named for *E. coli* YjiW. Many of the proteins appear in a conserved gene arrangement with a repressor resembling the Cro protein of bacteriophage, a DnaG-like primase, and an integrase/recombinase, suggesting that they are part of a prophage or mobile element. Particularly noteworthy is the fact that in several such gene clusters of *Erwinia carotovora*, the AbrB-like protein is preceeded by a related protein, which contains two AbrB-like sequences in a single chain (like MraZ) and includes a helical hairpin between copies, but lacks the first β strand of the second copy (Figure 2). This is the only group that combines homodimeric and single-chain versions of AbrB.

Hypothetical Proteins of Cyanobacteria

This last major group includes only hypothetical proteins of cyanobacteria and appears chromosomally linked to cytochrome C subunits and riboflavin synthase.

In addition to these groups, the cluster map contains a few sequences that are not assignable and have a peculiar domain structure (red dots in Figure 1). These are: (i) two RecG homologs from *Geobacter* spp., one found close to PrlF, and the other, connected to it, outside the Vir cluster; (ii) two permeases of Lactobacilli, one outside the PhoU cluster, and the other at the upper edge of the map (two other homologs in the sequence database were not captured by our search); and (iii) a protein from the crenarchaeon *Aeropyrum pernix* resembling acetolactate synthase small subunit, found between AbrB and MraZ-N (more distant homologs with the same domain structure in *Thermococcus* and *Pyrococcus* spp. were not captured by our search). In all of these proteins, the AbrB-like domain is at the N terminus (a full list of the proteins found in this study is given in the Supplemental Data available with this article online).

The size and spectrum of the protein families identified in this study show that AbrB-like proteins form a major group of prokaryotic transcription factors involved in diverse processes. The definition of this superfamily permits functional inferences for families with unknown biological roles, such as a transcriptional regulatory activity for MraZ. In addition, the study provides further structural information on AbrB-N by revealing the existence of two homologs with known crystal structures. The concordance of these two structures and their difference to that published for AbrB-N (1EKT) prompted us to reinvestigate the structure of AbrB-N by using NMR spectroscopy.

Structure Determination of AbrB-N

Resonance assignments for AbrB could be transferred almost completely from those previously published (Vaughn et al., 2000), with the exception of the termini, in which the sequences of the two constructs differ slightly. Analysis of backbone chemical shifts with the program TALOS (Cornilescu et al., 1999) indicated that the AbrB monomer contains four β strands, in contrast to the three previously reported, and one α helix (I22–L28). The additional strand (β2, G16–I20) showed sev-

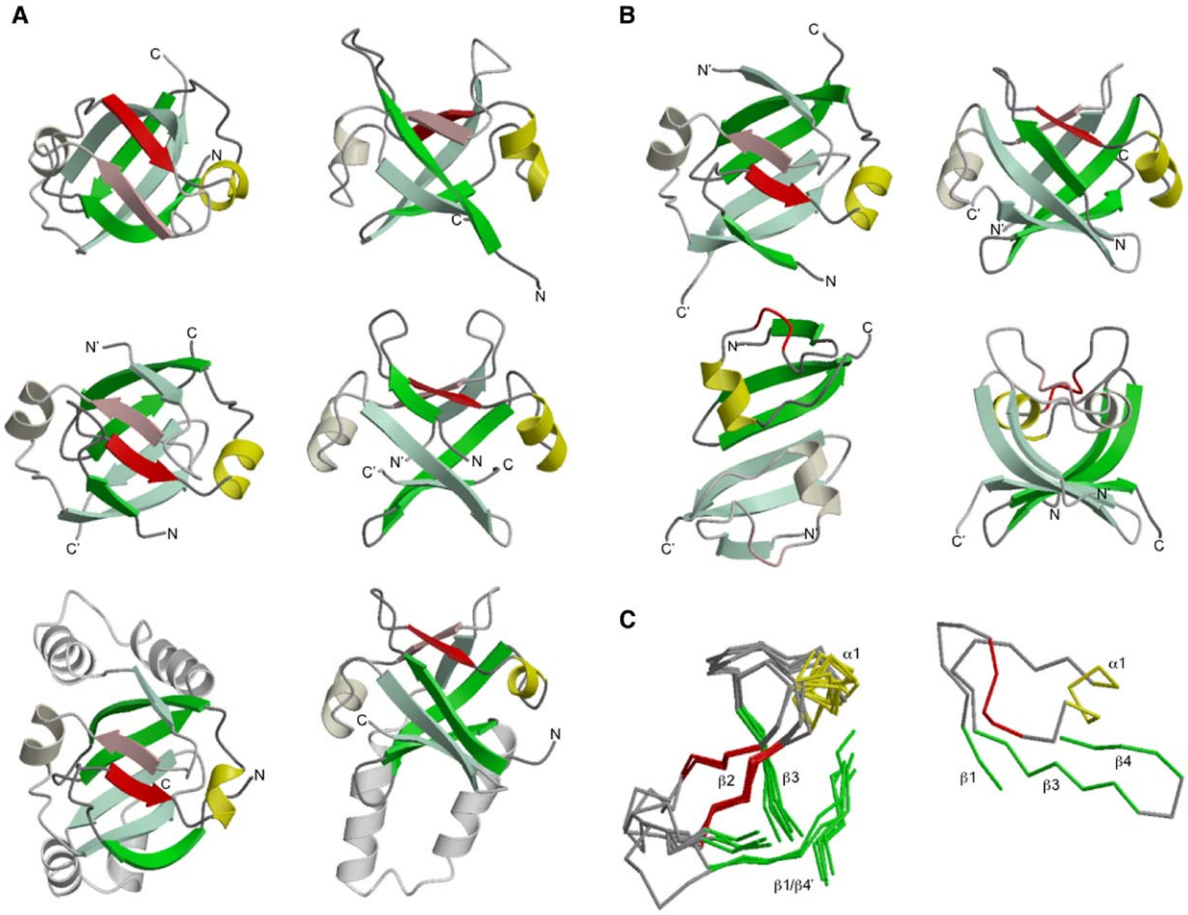

Figure 3. Structures of AbrB-N and Related β Barrels

(A) The double-psi barrel fold of VatN-N (1CZ4, top) and the swapped-hairpin barrels of MazE (1UB4, middle) and MraZ (1N0G, bottom). The top view (left) is related to the side view (right) by a 90° rotation around the horizontal axis. The secondary structure of one symmetrical half of each domain is shown in bold color, with the β2 strand in red. The swapping of position of this strand and the two additional β strands in the swapped-hairpin barrel are the main topological differences in the folds. VatN-N and MraZ are monomers, and MazE is a homodimer. For reasons of topology, the two (pseudo) symmetrical halves of MraZ must be joined by a helical hairpin (the hairpin appears after each half, but the C-terminal hairpin has been omitted from the side view for clarity). The typical "horned" profile provided by the β1–β2 loop is a striking feature of all three domains.

(B) Two structures for AbrB-N: the swapped-hairpin fold solved here (1YFB, top), and the looped-hinge helix fold previously reported by Vaughn et al. (2000) (1EKT, bottom). The views and coloring are as for (A). The previously reported structure cannot explain the formation of intermolecular β2–β2′ contacts, while the β1–β3/β1′–β3′ contacts are incorrectly assigned as intramolecular. The similarity of the structure presented here to MazE and MraZ is apparent.

(C) The conserved structural core of the swapped-hairpin and double-psi barrels. The left view shows a superposition of AbrB-N, MazE, and of both halves of MraZ and VatN-N. The right view shows the looped-hinge helix structure for AbrB-N, such that orientation of the helix and the following GD box are the same as in the left panel. The superposition shows the structural equivalence of β1 of the double-psi barrel with β4 of the swapped-hairpin barrel. A circular permutation cannot, however, be concluded, since the equivalence is to β4 of the symmetry-related subunit.

eral NOESY connectivities that were best explained by an intermolecular antiparallel β sheet contact with its symmetry-related equivalent, β2′. This observation is incompatible with the fold originally reported for the protein. The intermolecular contacts expected for this topology, e.g., a strong H^{α}-H^{α} contact between R17 and V19, were confirmed in filtered/edited NOESY experiments on a sample containing asymmetrically ^{15}N- and ^{13}C-labeled monomers (Figure 4). Intermolecular contacts were also observed between β1 (I8–V12) and β3′ (D34–D41). This is also inconsistent with the originally reported fold, in which the β1 strand made intramolecular contacts to β3. The contacts for β4 (K44–Y50), i.e., the intramolecular β3–β4 and intermolecular β4–β4′ contacts, were as previously reported. The overall fold was thus consistent with that of the identified homologs MazE and MraZ and rationalized the sequence conservation patterns of the AbrB superfamily (Figure 2).

Structural data for AbrB consisted of distance data derived from a 3D ^{15}N-HSQC-NOESY as well as several 2D NOESY spectra, $^{3}J_{HNH\alpha}$ coupling constants derived from an HNHA experiment, and chemical shift-derived backbone torsion angle restraints. An initial model was created by using all available data and was used as the

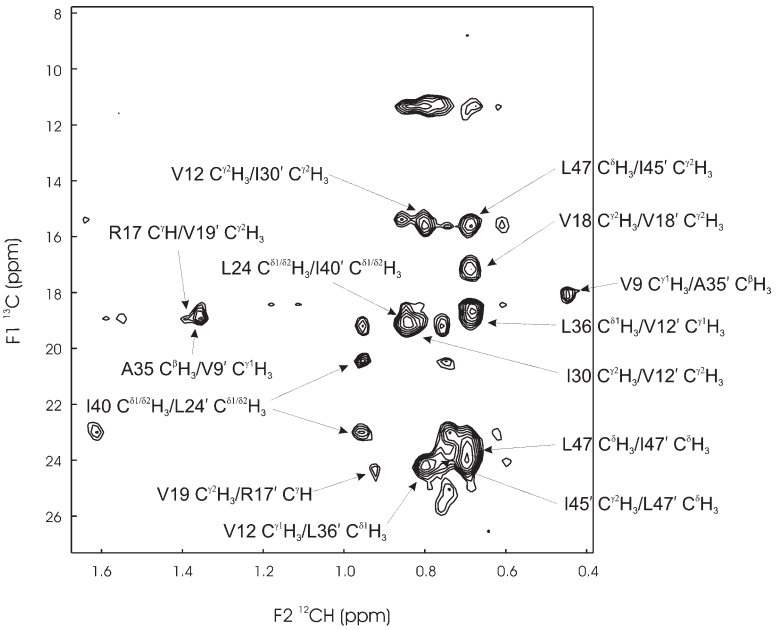

Figure 4. Intermolecular Contacts Define the AbrB-N Topology

A section of the $^{14}N,^{12}C$-filtered/^{13}C-edited 2D-NOESY spectrum containing primarily methyl-methyl contacts between hydrophobic residues is shown. Selected pairs of diagonally related crosspeaks are labeled. They represent contacts between β2 and β2′ (R17/V18–V18′/V19′), β1 and β3′ (V9/V12–A35′/L36′), β4 and β4′ (I45/L47–L47′), and between the α helix (L24) and the GD box (I30) and the hydrophobic core of the protein. A total of 22 crosspeaks could be assigned in this spectrum, contributing to the total of 55 intermolecular distance restraints included in calculations.

starting point for iterative assignment of further NOE connectivities, resulting in the final set of experimental restraints described in Table 1. The structure ensemble is well defined (Figure 5), with an rmsd for the final set of 20 dimeric structures (residues T6–Y50) of 0.20 Å for backbone atoms and 0.70 Å for all heavy atoms.

The monomer structure (Figure 5) can be described as a pair of β hairpins linked by the α helix and the GD box. The monomer does not contain an independent hydrophobic core, and it thus requires dimerization to form a viable fold. This dimerization takes place by interleaving the four β hairpin elements, such that each only makes contacts to those of the dimeric partner. The interleaving is ensured by the extended nature of the connector between the α helix and the GD motif, in which the two large hydrophobic residues L28 and I30 are buried in the opening of the barrel, providing for the rigidity of the crossover connection. The result is an eight-stranded, swapped-hairpin β barrel. The two helices close the barrel at each end. The two GD boxes have a key structural role, forming β turns that are anchored into the core of the barrel by flanking hydrophobic residues. Hydrogen bonds between the turn and the β1 strand of the dimeric partner fix the positions of the β1–β2 loops. These loops project above the surface of the barrel, giving the protein its characteristic horned profile (Figure 3) and forming a cleft rich in positive charge that, in MazE, has been implicated in interaction with DNA (Loris et al., 2003).

The fold presented here is markedly different from that of Vaughn et al. (2000), in which β2, and therefore the β1–β2 hairpin, was absent and no interleaving of monomer elements occurred (Figure 3). In the resulting structure, each monomer was an independent folding unit, and dimerization took place in a simple side-by-side manner. The formal discrimination of inter- and intramolecular NOESY contacts in the current study with dimer samples with asymmetrically labeled monomers refutes this model.

Comparison of the AbrB structure presented here to those of the identified homologs, MraZ (1MVF) and MazE (1N0G), shows the proteins to share a common fold (Figure 3) while revealing differences in structural details. Both MraZ and AbrB have short β1–β2 loops, comprised completely of a β turn (D13–G16 in AbrB) and characterized by a web of interactions between the side chains of an aspartic acid and an arginine residue (D13 and R17 in AbrB) and neighboring backbone amides (the backbone rmsd of AbrB V12–R17 to 1N0G T34–R40 is 0.4 Å). AbrB and MraZ lack formal β sheet contacts in the β1–β2 hairpin, meaning that they must be considered pseudobarrels. In contrast, the β1–β2 loop of MazE is somewhat longer, and a pair of hydrogen bond contacts between β1 and β2 closes the barrel. AbrB shares with MazE the β turn structure over the GD box residues typical of the fold family, while MraZ lacks both of these residues and the β turn. A conserved PxxxR sequence within the α1 helix is also typical of the family (Figure 2). In AbrB and all other structures in which this motif is present, the arginine forms a side chain H bond to the backbone carbonyl immediately preceding the proline, thus facilitating the sharp change in chain direction between β2 and α1. This motif is only present in the first half of MraZ, while, in the second half and in MazE, a large hydrophobic residue replaces the arginine (I and M, respectively). Thus, many of the conserved structural features do not seem to be determinants of the overall fold, explaining the imperfect conservation pattern of the involved residues in the AbrB superfamily.

Discussion

In the process of evaluating the discrepancy between an evolutionary scenario for the origin of double-psi β barrels and the experimentally determined fold for AbrB-N, we defined a new superfamily of prokaryotic

Table 1. Structural Statistics and Atomic Rms Deviations

Structural Statistics[a]		
	SA	<SA>$_r$
Rmsd from distance restraints (Å)[b]		
All (418)	0.027 ± 0.0010	0.026
Intraresidue (145)	0.012 ± 0.0031	0.012
Interresidue sequential (152)	0.026 ± 0.0016	0.024
Medium range (37)	0.036 ± 0.0013	0.035
Long range (29)	0.047 ± 0.0059	0.048
Intermolecular (55)	0.033 ± 0.0013	0.032
Rmsd from dihedral restraints (106)	0.463 ± 0.073	0.427
Rmsd from J coupling restraints (Hz) (32)	0.568 ± 0.012	0.586
H bond restraint violations (Å/°)[c] (35)	2.10 ± 0.19/16.8 ± 8.6	2.11/14.1
Deviations from ideal covalent geometry		
Bonds (Å × 10^{-3})	1.67 ± 0.032	1.63
Angles (°)	0.502 ± 0.003	0.497
Impropers (°)	1.272 ± 0.051	1.200
Structure quality indicators[d]		
Ramachandran map regions (%)	88.9/10.6/0.5/0.0	90.2/9.8/0./ 0.0
Bad contacts per 100 residues	7.6 ± 2.4	7.5

Atomic Rms Differences (Å)[e]	SA versus <SA>		SA versus <SA>$_r$	
	Backbone	All	Backbone	All
All residues	1.83 ± 0.78	2.31 ± 0.713	2.17 ± 1.262	2.82 ± 1.131
Ordered residues[f] (dimer)	0.20 ± 0.065	0.70 ± 0.070	0.27 ± 0.067	0.98 ± 0.121
Ordered residues (monomer)	0.19 ± 0.064	0.70 ± 0.070	0.26 ± 0.066	0.98 ± 0.123
<SA> versus <SA>$_r$[g]	0.18	0.69		

[a] Structures are labeled as follows: SA, the set of 25 final simulated annealing structures; <SA>, the mean structure calculated by averaging the coordinates of SA structures after fitting over secondary structure elements; <SA>$_r$, the structure obtained by regularizing the mean structure under experimental restraints.
[b] Numbers in brackets indicate the number of restraints of each type per monomer.
[c] H bonds were restrained by treating them as pseudocovalent bonds (see Experimental Procedures). Deviations are expressed as the average distance/average deviation from linearity for restrained H bonds.
[d] Determined by using the program PROCHECK (Laskowski et al., 1993). Percentages are for residues in allowed/additionally allowed/generously allowed/disallowed regions of the Ramachandran map.
[e] Based on heavy atoms superimpositions.
[f] Defined as residues T6–Y50.
[g] Rms difference for superimposition over ordered residues.

transcription factors, which includes transition-state regulators, the antitoxins of several classes of postsegregational killing systems, putative regulators of cell wall biosynthesis, regulators of phosphate uptake, and a large number of proteins of as yet unknown activity (Figures 1 and 2). This superfamily contained two members with known crystal structure that resembled each other and differed from the solution structure reported for AbrB-N, prompting us to redetermine its structure. Our results are in agreement with the crystal structures (Figure 3); AbrB-N consists of four β strands arranged into two harpins, which are interleaved in the dimer (Figure 5), leading us to name this fold the swapped-hairpin barrel. In side view, the two N-terminal hairpins curve upward, giving the barrel a characteristic horned appearance and forming a deep binding cleft.

The similarities of the double-psi barrel to the swapped-hairpin barrel are striking (Figure 3). Despite a different overall fold, many structural features of VatN-N remarkably resemble AbrB-N (Figure 3C). Particularly similar is the turn over the GD box and its hydrogen bonding interaction with the symmetry-related β1 strand. The arginine of the PxxxR motif also makes the expected side chain-backbone hydrogen bond contacts (although the proline has been replaced by an aspartic acid residue in both VatN-N repeats). The psi loops curve upward in the same way as the β1–β2 loops of the swapped-hairpin barrel, giving the two folds a similar profile, and forming a deep, positively charged cleft. Indeed, this cleft has been implicated in substrate interactions of VatN (Coles et al., 1999), and, furthermore, expression of VatN-N alone leads to spontaneous dimerization and to a low but measurable DNA binding activity, as detected in band-shift assays with heterologous DNA from *E. coli* (data not shown). All of these observations support a homologous origin for the two folds, which we propose to denote as cradle-loop barrels, in view of their peculiar profile and cradle-shaped binding surface. We have previously hypothesized that folded protein domains arose through the fusion and recombination of a smaller number of subdomain-sized peptides (antecedent domain segments) (Lupas et al., 2001; Soding and Lupas, 2003), which themselves emerged in the context of RNA-based replication and catalysis (the "RNA world"). The strong similarity between the swapped-hairpin and double-psi barrels in the core βαβ region suggests that this element might correspond to such an antecedent domain segment.

Nevertheless, the topological relationship between the two folds is not obvious. The structural model we

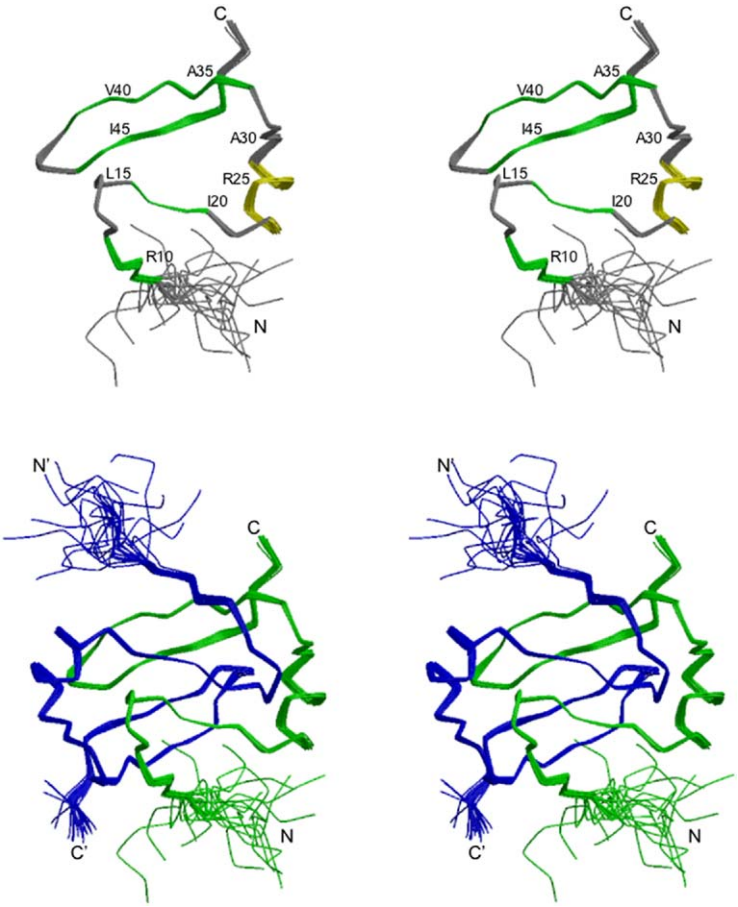

Figure 5. A Stereo View of the Ensemble of the 25 Lowest-Energy Structures for AbrB-N

The upper panel shows the AbrB-N monomer colored by secondary structure, while the lower panel shows the dimer colored by monomer. The superimposition is over residues T7–Y50 of each monomer, resulting in an rmsd of 0.20 Å for backbone atoms and 0.70 Å for all heavy atoms.

originally proposed for AbrB (Coles et al., 1999) by extrapolation from VatN-N is incorrect in that the barrel is eight-, not six-stranded, and the two subunits forming the barrel are interleaved, not face-to-face. Although the position of β1 in each repeat of the double-psi barrel is equivalent to that of β4 in the swapped-hairpin barrel (Figure 3C), in accordance with our prediction, the relationship holds for β1 and β4′ (i.e., β4 of the symmetry-related half), introducing an unanticipated topological complexity and essentially ruling out a transformation by circular permutation, as we originally proposed. Thus, although the swapped-hairpin barrel is topologically simpler than the double-psi barrel by virtue of lacking the knotted psi loops, it does not seem likely to represent a precursor form. Rather, both barrel types appear to have evolved from a yet simpler third type. We are currently investigating the structures of several sequences with properties intermediate between AbrB-N and VatN-N in order to identify this third type.

Experimental Procedures

Sequence Searches and Cluster Analysis

For sequence searches on the nonredundant protein sequence database, we developed a method, HHsenser, by combining the search strategy encoded in SENSER (Koretke et al., 2002) with a method for comparing profile Hidden Markov Models (HHsearch) (Soding, 2004). HHsenser is thus an itermediate profile search method that is based on HMM-HMM comparison. HHsenser first employs PSI-Blast (Altschul et al., 1997) to generate an initial alignment of relatively close homologs of the query sequence (AbrB, in this case). It then selects representative sequences (maximum pairwise sequence identity of 40% and PSI-Blast E values of up to 1) in the vicinity of the query sequence as seeds for building new alignments with PSI-Blast. HMM-HMM comparison by HHsearch is then employed to check if the new alignment is homologous to the superalignment of all accepted homologs of the query sequence. If the alignment is accepted, it is included in the superalignment. In addition, the representative sequences in the vicinity of the newly accepted alignment are added to the list of seed sequences, and the process is continued until all seeds have been processed. A series of heuristics ensures that run time is kept to a minimum by avoiding PSI-Blast searches that will probably not result in a homologous alignment. The AbrB superfamily alignment was generated in 8.5 hr on a single 3 GHz 64-bit AMD CPU.

After convergence of the sequence searches, the obtained sequences were clustered by using the program CLANS (Frickey and Lupas, 2004) at a P-value cutoff of e-4. The sequences corresponded in all cases to the parts identified as being similar to AbrB and were thus almost always fragments of the complete sequences. In some proteins, two fragments homologous to AbrB were recovered. The resulting map is shown in Figure 1. We reexamined the individual clusters manually, as well as all proteins not clearly assigned to a cluster, but we did not detect any false positives. We found, though, that the search routine had missed a small number of sequences in each cluster, which could be identified by taking that cluster as the starting point. We are not certain at present of the reason. We also observed that the incorrect assignment of start codons in the database had shortened the AbrB domains in some proteins, causing them to become outliers to their clusters. This is seen particularly clearly in the MraZ clusters, where the N

domain cluster is much more irregular than the C domain cluster. The most extreme example of a miscalled start codon was in *Geobacter* RecG, where the annotated *G. sulfurreducens* homolog lacks practically the entire AbrB domain and thus groups well away from *G. metalloreducens* RecG, even though the two domains actually share 50% sequence identity when the correct start codon is used.

Sample Preparation

The AbrB-N construct (encoding amino residues 1–53 of AbrB, gi113009) was amplified from *B. subtilis* PY79 chromosomal DNA by polymerase chain reaction (PCR) and was cloned into the pet30b vector (Novagen). The construct contained a His6-tag at the amino terminus to facilitate purification. For expression in *E. coli*, cells were grown in LB medium at 37°C, induced at an OD_{600} of ~0.6 with 1 mM IPTG, and was harvested after 4 hr. Uniformly ^{15}N- or ^{13}C-labeled AbrB-N was made by growing bacteria in M9 minimal medium by using $^{15}NH_4Cl$ (0.7 g/l) and $^{13}C_6$-glucose (2 g/l) as sole nitrogen or carbon sources. An asymmetrically ^{15}N- and ^{13}C-labeled AbrB-N sample was made by combining the same amounts of harvested cells prior to lysis. Proteins were purified by a combination of immobilized metal affinity chromatography (IMAC), ion exchange, and gel sizing chromatography. Purification under denaturing conditions allowed for statistical mixing of the labeled components, which was later confirmed by NMR. For NMR measurements, samples were concentrated to 8 mg/ml in buffer containing 20 mM potassium phosphate, 50 mM KCl, 0.02% (w/v) NaN_3 (pH 5.8).

NMR Spectroscopy

All spectra were recorded at 305 K on Bruker DMX600, DMX750, and DMX900 spectrometers. Resonance assignments were taken from Vaughn et al. (2000) and were confirmed either through NOESY connectivities (for sequential assignments) or in a CCH-COSY experiment on the asymmetrically labeled sample (for side chain assignments). This sample was additionally used to obtain carbonyl assignments for 36 of 52 residues by using a HACACO experiment. The stereospecific assignments of prochiral groups and the resulting rotamer assignments were also checked.

Distance data were derived from a 3D-^{15}N-HSQC-NOESY on a ^{15}N-labeled sample and a 2D-NOESY spectrum recorded on an unlabeled sample. Intermolecular NOESY contacts were identified in the asymmetrically ^{15}N- and ^{13}C-labeled sample by using a $^{14}N,^{12}C$-filtered/^{13}C-edited 2D-NOESY spectrum (see Figure 4). Residual diagonal signals in this spectrum were effectively eliminated by subtraction of an identical spectrum run with minimal NOESY mixing time. Intramolecular $^{14}NH/^{13}CH$ contacts were exclusively observable in this spectrum and were thus discriminated from intermolecular $^{15}NH/^{13}CH$ contacts. Contacts identified in the filtered/edited spectrum were quantified when possible in other spectra.

NOESY crosspeaks in the 3D spectra were converted into distance ranges after rescaling according to corresponding HSQC intensities. Crosspeaks were divided into four classes: strong, medium, weak, and very weak, which resulted in restraints on upper distances of 2.7, 3.2, 4.0, and 5.0 Å, respectively. Lower distance restraints were also included for very weak or absent sequential H^N-H^N crosspeaks by using a minimum distance of 3.2 Å and medium intensity or weaker sequential and intraresidue H^N-H^α crosspeaks by using a minimum distance of 2.7 Å. Allowances for the use of pseudoatoms (using r^{-6} averaging) were added for methyl groups and nonstereospecifically assigned methylene groups. Dihedral angle restraints were derived for backbone ϕ and ψ angles based on C^α, C^β, and H^α chemical shifts by using the program TALOS (Cornilescu et al., 1999). Restraints were applied for the 35 high-confidence predictions found by the program by using the calculated range ±10°. In addition, direct coupling constant restraints were included for the backbone ϕ angles of 32 residues based on $^3J_{HNH\alpha}$ coupling constants measured from an HNHA experiment. Hydrogen bond restraints were applied for 26 residues in secondary structure with low water exchange rates, as judged by the strength of water exchange crosspeaks in the ^{15}N-HSQC-NOESY spectrum, and where donor-acceptor pairs were consistently identified in preliminary calculations. This included 11 intermolecular hydrogen bonds. The restraints were applied via inclusion of pseudocovalent bonds as described by Truffault et al. (2001).

Structures were calculated with XPLOR (NIH version 2.9.3) by using standard protocols. Experimental restraints were applied only to one monomer, with noncrystallographic symmetry restraints over the backbone of ordered residues (T7–Y50) used to ensure the symmetry of the dimer. Sets of 50 structures were calculated, and a final set of 25 was chosen on the basis of lowest restraint violations. An average structure was calculated and regularized to give a structure representative of the ensemble.

Supplemental Data

Supplemental Data including a full list of proteins used to construct the cluster map shown in Figure 1 are available at http://www.structure.org/cgi/content/full/13/6/919/DC1.

Acknowledgments

The authors thank Prof. Horst Kessler and the staff of the Bavarian Nuclear Magnetic Resonance (NMR) Centre at the Technical University, Munich for access to spectrometers and technical support. The bioinformatics study was conducted by A.N.L., J.S., T.F., and K.K., with support from S.D. The biochemistry is the work of S.D., with support from J.M. The NMR structure was determined by M.C. and V.T. The authors declare that they have no competing financial interests.

Received: February 25, 2005
Revised: March 29, 2005
Accepted: March 29, 2005
Published: June 7, 2005

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Anantharaman, V., and Aravind, L. (2003). New connections in the prokaryotic toxin-antitoxin network: relationship with the eukaryotic nonsense-mediated RNA decay system. Genome Biol. *4*, R81.

Andreeva, A., Howorth, D., Brenner, S.E., Hubbard, T.J., Chothia, C., and Murzin, A.G. (2004). SCOP database in 2004: refinements integrate structure and sequence family data. Nucleic Acids Res. *32*, D226–D229.

Bagyan, I., Hobot, J., and Cutting, S. (1996). A compartmentalized regulator of developmental gene expression in *Bacillus subtilis*. J. Bacteriol. *178*, 4500–4507.

Baird, L., and Georgopoulos, C. (1990). Identification, cloning, and characterization of the *Escherichia coli* sohA gene, a suppressor of the htrA (degP) null phenotype. J. Bacteriol. *172*, 1587–1594.

Benson, L.M., Vaughn, J.L., Strauch, M.A., Bobay, B.G., Thompson, R., Naylor, S., and 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.

Bobay, B.G., Benson, L., Naylor, S., Feeney, B., Clark, A.C., Goshe, M.B., Strauch, M.A., Thompson, R., and Cavanagh, J. (2004). Evaluation of the DNA binding tendencies of the transition state regulator AbrB. Biochemistry *43*, 16106–16118.

Castillo, R.M., Mizuguchi, K., Dhanaraj, V., Albert, A., Blundell, T.L., and Murzin, A.G. (1999). A six-stranded double-psi beta barrel is shared by several protein superfamilies. Struct. Fold. Des. *7*, 227–236.

Cavanagh, J., Thompson, R., Bobay, B., Benson, L.M., and Naylor, S. (2002). Stoichiometries of protein-protein/DNA binding and con-

formational changes for the transition-state regulator AbrB measured by pseudo cell-size exclusion chromatography-mass spectrometry. Biochemistry *41*, 7859–7865.

Chen, S., Jancrick, J., Yokota, H., Kim, R., and Kim, S.H. (2004). Crystal structure of a protein associated with cell division from *Mycoplasma pneumoniae* (GI: 13508053): a novel fold with a conserved sequence motif. Proteins *55*, 785–791.

Coles, M., Diercks, T., Liermann, J., Groger, A., Rockel, B., Baumeister, W., Koretke, K.K., Lupas, A., Peters, J., and Kessler, H. (1999). The solution structure of VAT-N reveals a 'missing link' in the evolution of complex enzymes from a simple betaalphabetabeta element. Curr. Biol. *9*, 1158–1168.

Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biomol. NMR *13*, 289–302.

Dong, T.C., Cutting, S.M., and Lewis, R.J. (2004). DNA-binding studies on the *Bacillus subtilis* transcriptional regulator and AbrB homologue, SpoVT. FEMS Microbiol. Lett. *233*, 247–256.

Engelberg-Kulka, H., and Glaser, G. (1999). Addiction modules and programmed cell death and antideath in bacterial cultures. Annu. Rev. Microbiol. *53*, 43–70.

Frickey, T., and Lupas, A. (2004). CLANS: a Java application for visualizing protein families based on pairwise similarity. Bioinformatics *20*, 3702–3704.

Gerdes, K. (2000). Toxin-antitoxin modules may regulate synthesis of macromolecules during nutritional stress. J. Bacteriol. *182*, 561–572.

Gotfredsen, M., and Gerdes, K. (1998). The *Escherichia coli* relBE genes belong to a new toxin-antitoxin gene family. Mol. Microbiol. *29*, 1065–1076.

Grishin, N.V. (2001). Fold change in evolution of protein structures. J. Struct. Biol. *134*, 167–185.

Hamon, M.A., Stanley, N.R., Britton, R.A., Grossman, A.D., and Lazazzera, B.A. (2004). Identification of AbrB-regulated genes involved in biofilm formation by *Bacillus subtilis*. Mol. Microbiol. *52*, 847–860.

Kamada, K., Hanaoka, F., and Burley, S.K. (2003). Crystal structure of the MazE/MazF complex: molecular bases of antidote-toxin recognition. Mol. Cell *11*, 875–884.

Katz, M.E., Strugnell, R.A., and Rood, J.I. (1992). Molecular characterization of a genomic region associated with virulence in *Dichelobacter nodosus*. Infect. Immun. *60*, 4586–4592.

Klein, W., and Marahiel, M.A. (2002). Structure-function relationship and regulation of two *Bacillus subtilis* DNA-binding proteins, HBsu and AbrB. J. Mol. Microbiol. Biotechnol. *4*, 323–329.

Koretke, K.K., Russell, R.B., and Lupas, A.N. (2002). Fold recognition without folds. Protein Sci. *11*, 1575–1579.

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

Lehnherr, H., Maguin, E., Jafri, S., and Yarmolinsky, M.B. (1993). Plasmid addiction genes of bacteriophage P1: doc, which causes cell death on curing of prophage, and phd, which prevents host death when prophage is retained. J. Mol. Biol. *233*, 414–428.

Loris, R., Marianovsky, I., Lah, J., Laeremans, T., Engelberg-Kulka, H., Glaser, G., Muyldermans, S., and Wyns, L. (2003). Crystal structure of the intrinsically flexible addiction antidote MazE. J. Biol. Chem. *278*, 28252–28257.

Lupas, A.N., Ponting, C.P., and Russell, R.B. (2001). On the evolution of protein folds: are similar motifs in different protein folds the result of convergence, insertion, or relics of an ancient peptide world? J. Struct. Biol. *134*, 191–203.

Mayer, M.P., Bueno, L.C., Hansen, E.J., and DiRienzo, J.M. (1999). Identification of a cytolethal distending toxin gene locus and features of a virulence-associated region in Actinobacillus actinomycetemcomitans. Infect. Immun. *67*, 1227–1237.

Pamnani, V., Tamura, T., Lupas, A., Peters, J., Cejka, Z., Ashraf, W., and Baumeister, W. (1997). Cloning, sequencing and expression of VAT, a CDC48/p97 ATPase homologue from the archaeon Thermoplasma acidophilum. FEBS Lett. *404*, 263–268.

Phillips, Z.E., and Strauch, M.A. (2002). *Bacillus subtilis* sporulation and stationary phase gene expression. Cell. Mol. Life Sci. *59*, 392–402.

Pullinger, G.D., and Lax, A.J. (1992). A *Salmonella dublin* virulence plasmid locus that affects bacterial growth under nutrient-limited conditions. Mol. Microbiol. *6*, 1631–1643.

Radnedge, L., Davis, M.A., Youngren, B., and Austin, S.J. (1997). Plasmid maintenance functions of the large virulence plasmid of *Shigella flexneri*. J. Bacteriol. *179*, 3670–3675.

Shafikhani, S.H., and Leighton, T. (2004). AbrB and Spo0E control the proper timing of sporulation in *Bacillus subtilis*. Curr. Microbiol. *48*, 262–269.

Snyder, W.B., and Silhavy, T.J. (1992). Enhanced export of beta-galactosidase fusion proteins in prlF mutants is Lon dependent. J. Bacteriol. *174*, 5661–5668.

Soding, J. (2004). Protein homology detection by HMM-HMM comparison. Bioinformatics *21*, 951–960.

Soding, J., and Lupas, A.N. (2003). More than the sum of their parts: on the evolution of proteins from peptides. Bioessays *25*, 837–846.

Strauch, M.A. (1995). Delineation of AbrB-binding sites on the *Bacillus subtilis* spo0H, kinB, ftsAZ, and pbpE promoters and use of a derived homology to identify a previously unsuspected binding site in the bsuB1 methylase promoter. J. Bacteriol. *177*, 6999–7002.

Strauch, M.A., and Hoch, J.A. (1993). Transition-state regulators: sentinels of *Bacillus subtilis* post-exponential gene expression. Mol. Microbiol. *7*, 337–342.

Strauch, M.A., Spiegelman, G.B., Perego, M., Johnson, W.C., Burbulys, D., and Hoch, J.A. (1989). The transition state transcription regulator abrB of *Bacillus subtilis* is a DNA binding protein. EMBO J. *8*, 1615–1621.

Truffault, V., Coles, M., Diercks, T., Abelmann, K., Eberhardt, S., Luttgen, H., Bacher, A., and Kessler, H. (2001). The solution structure of the N-terminal domain of riboflavin synthase. J. Mol. Biol. *309*, 949–960.

Vaughn, J.L., Feher, V., Naylor, S., Strauch, M.A., and Cavanagh, J. (2000). Novel DNA binding domain and genetic regulation model of *Bacillus subtilis* transition state regulator abrB. Nat. Struct. Biol. *7*, 1139–1146.

Vaughn, J.L., Feher, V.A., Bracken, C., and Cavanagh, J. (2001). The DNA-binding domain in the *Bacillus subtilis* transition-state regulator AbrB employs significant motion for promiscuous DNA recognition. J. Mol. Biol. *305*, 429–439.

Vicente, M., Gomez, M.J., and Ayala, J.A. (1998). Regulation of transcription of cell division genes in the *Escherichia coli* dcw cluster. Cell. Mol. Life Sci. *54*, 317–324.

Wanner, B.L. (1993). Gene regulation by phosphate in enteric bacteria. J. Cell. Biochem. *51*, 47–54.

Xu, K., and Strauch, M.A. (2001). DNA-binding activity of amino-terminal domains of the *Bacillus subtilis* AbrB protein. J. Bacteriol. *183*, 4094–4098.

Zhang, Y., Zhang, J., Hoeflich, K.P., Ikura, M., Qing, G., and Inouye, M. (2003). MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in *Escherichia coli*. Mol. Cell *12*, 913–923.

Zhang, J., Zhang, Y., Zhu, L., Suzuki, M., and Inouye, M. (2004). Interference of mRNA function by sequence-specific endoribonuclease PemK. J. Biol. Chem. *279*, 20678–20684.

Accession Numbers

The coordinates for the ensembles and the regularized average structure have been deposited in the Protein Data Bank (accession codes 1YSF and 1YFB, respectively).

Note Added in Proof

During processing of this manuscript, a corrigendum retracting the PDB entry 1EKT has appeared in press (Vaughn et al. [2005], Nat. Mol. Struct. Biol. *12*, 380), and a revised PDB entry, 1Z0R, was submitted.