Published in final edited form as:

J Mol Biol. 2007 January 26; 365(4): 1130–1142. doi:10.1016/j.jmb.2006.10.066.

Solution structure of *E.coli* PapI, a key regulator of the pap pili phase variation

Tetsuya Kawamura, Lisa Uyen K. Le, Hongjun Zhou, and Frederick W. Dahlquist[¥]

Department of Chemistry and Biochemistry, University of California Santa Barbara, Santa Barbara, CA, 93106-9510

Abstract

Pylonephritis-associated pili (pap) allow uropathogenic *Escherichia coli* to bind to epithelial cells and play an important role in urinary tract infection. Expression of pap pili is controlled by a phase-variation mechanism, based on the two distinct heritable states that are the result of adenine N6-methylation in either of the two GATC sequences in its regulatory region. Methylation status of these two sequences is sensed by the action of two proteins, Lrp and PapI, and they play a central role in determining pap pili gene expression in both phase-ON and phase-OFF cells. We used modern nuclear magnetic resonance techniques to determine the solution structure and backbone dynamics of PapI. We found its overall fold closely resembles that of the winged helix-turn-helix family of DNA-binding proteins. We also determined that PapI possesses its own DNA-binding activity, albeit non-sequence specific, independent of Lrp. PapI appears to bind to DNA with a K_d in the 10 μ M range. Possible mechanisms by which PapI might participate in the regulation of the *pap* operon are discussed in light of these new findings.

Keywords

Lrp; *pap* pili; phase variation; winged helix-turn-helix; DNA-binding protein

Introduction

The leucine-responsive regulatory protein (Lrp)-dependent epigenetic regulation of the *papBA* operon has been extensively studied, both genetically and biochemically.^{1,2} Modulation of *pap* expression involves a phase variation mechanism based on the status of the two dam methylation sites (GATC) in its promoter region (Figure 1). The *pap* promoter region contains six Lrp-binding sites, which can be divided into two functional clusters, proximal (sites 1, 2, and 3) and distal (sites 4, 5, and 6) to the transcription start. In phase-OFF cells, Lrp occupies the cluster overlapping the proximal GATC site (“GATC^{prox}”). This blocks the methylation of GATC^{prox} and inhibits the transcription start probably by promoter occlusion. In phase-ON cells, Lrp binds to the distal site cluster overlapping the other dam methylation site (“GATC^{dist}”), blocking its methylation, leaving GATC^{prox} open for methylation, and allowing the transcription to start.^{3,4,5} This modulation in the DNA-binding activity of Lrp plays a central role in the *pap* phase variation, and, as evidenced by early genetic studies, it depends on the local regulator PapI.^{6,7,8} The entire mechanism of *pap* regulation involves

[¥]corresponding author, Email: dahlquist@chem.ucsb.edu, Phone: 805-893-5326, Fax: 805-893-4120.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

other proteins, but in vitro studies with isolated components showed that PapI directly interacts with Lrp and that formation of the Lrp-PapI complex is sufficient for the methylation-dependent affinity change to occur.⁹

Recent results from the Low laboratory have clearly demonstrated that the specificity of the PapI effect on the binding of Lrp to individual *pap* binding sites is centered on the two GATC sites located at subsites 2 and 5.⁹ The Low laboratory used competitive band shift assays to monitor the binding of Lrp to each of the six *pap* binding sites in the presence and absence of PapI. Their results showed that PapI substantially enhances the affinity of Lrp for sites 2 and 5 but has no significant effect for the other four subsites. In the presence of adenine N6-methylation of the GATC sequences in sites 2 and 5, the effect of PapI on the affinity of Lrp for these sites is essentially removed. Thus PapI seems to modulate the DNA-binding activity of Lrp by providing the bulk of the differential recognition of the methylation states of the GATC sites.

How PapI confers such ability to Lrp is not known. The possibility that PapI might be a DNA-binding protein of its own has been raised previously.^{7,9} PapI belongs to a family of proteins known as “FaeA-like” proteins (Pfam accession number PF04703).^{10,11} PF04703 consists of 12 other proteins (and their isoforms), all of which are known or proposed orthologs of PapI. While no structural studies have been done on these proteins, they are hypothesized to be winged helix-turn-helix DNA-binding proteins. However, previous attempts to detect the DNA-binding activity of PapI independent of Lrp have been inconclusive.^{7,9}

In this report, we present the solution structure and backbone dynamics of PapI. Our results show that PapI is indeed a winged helix-turn-helix protein. We also found that PapI is a non-specific DNA-binding protein, at least in the absence of Lrp, with a dissociation constant (K_d) in the 10 μ M range in the presence of 150 mM salt. Possible mechanisms by which PapI might participate in the regulation of the *pap* operon are discussed in light of these new findings.

Results

NMR Assignments of PapI

Of the two proposed PapI transcription starts, the downstream site (PapI-2) was chosen for the construct used in this study.⁸ PapI-1, the alternative product, is longer than PapI-2 by four N-terminal residues. We initially attempted to use the wild-type (un-tagged) PapI-2 for this study. In our hands, however, WT PapI-2 was highly prone to proteolysis, and it did not have the long-term stability necessary for the NMR assignments, although we were able to collect its ¹H-¹⁵N HSQC spectra.¹² We constructed a C-terminally his-tagged PapI, which could be more rigorously purified. WT PapI-2, his-tagged PapI-2, and his-tagged PapI-1 gave essentially identical ¹H-¹⁵N HSQC spectra, except for the peaks that belong to the terminal residues and a few adjacent residues affected presumably by their presence (data not shown). This leads us to believe that his-tagged PapI-1 and PapI-2 fold into a structure highly similar, if not identical, to that of the un-tagged WT protein. In this report, we refer to the his-tagged PapI-2 as PapI, unless otherwise specified, and all subsequent NMR work was done on this construct.

Nearly complete backbone- and sidechain assignments of PapI were made using 3D and 4D triple resonance experiments. Figure 2 shows the backbone amide assignment of PapI (81 residues). The backbone assignment was made from Gln3 through His77 (the second histidine of the six-histidine tag) except the lone proline (residue 48). The ¹H-¹⁵N correlation spectrum exhibits excellent dispersion, indicative of a well-defined structure. A number of intense peaks were observed between 8.0 ppm and 8.5 ppm region. Most of them were assigned to the C-terminal region of the protein.

Secondary and Tertiary Structure

PapI structure was calculated with the program X-PLOR¹³ using 1153 NOE constraints supplemented with dihedral angle- and hydrogen-bond (H-bond) constraints. Statistical evaluation of the resulting structure is summarized in Table 1. Figure 3a shows a superposition of the resulting 25 lowest-energy structures of the 50 total calculations. The average backbone rmsd for the structured regions (residues 2–10, 17–24, 28–40, 44–47, and 56–59) was 0.40 Å.

The structure (PDB ID: 2HTJ) shows a well-folded domain containing three helices packed against a two-stranded β -sheet (denoted as $\alpha 1$ - $\alpha 2$ - $\alpha 3$ - $\beta 1$ - $\beta 2$), followed by a highly flexible, 20 residue long C-terminal tail, which includes the 6-histidine-tag and 2 residues introduced during the cloning (Figure 3b). The secondary structure features are clearly shown in the C α chemical shift index (Figure 3c).¹⁴ We detected a number of sequential NOEs between residues within the three helices, consistent with the patterns expected from a helical structure.¹⁵ The effects of such inter-residue contacts are evident in the N-terminus of PapI, for it does not show signs of frailty often seen at the extremities of peptide chains, and $\alpha 1$ seems to start at the beginning of the protein. The distinct NOE pattern indicative of antiparallel β -sheets was observed for residues between the two β -strands made of residues 44–47 and 56–59 (Figure 3d). We also observed a network of NOEs between the core residues contributed by the three helices and the β -sheet, indicating the formation of a tight compact structure.

As for the residues not engaged in the conventional secondary structures, there appears to be appreciable rigidity to the loops that follow α -helices 1, 2, and 3. We detected NOEs between these loops and surrounding regions. Not surprisingly, rmsd for the backbone atoms in these loops was noticeably lower than that for other residues not involved in α -helices or β -sheets (Figure 4d). A small patch of markedly weak NOEs forming a pattern closely resembling that of antiparallel β -sheet was detected between the two residues (Gly15 and Lys16) immediately preceding $\alpha 2$ and a part of $\beta 2$ (Tyr56 and Trp57, broken arrows in Figure 3d). Our calculations indicated that their dihedral angles in the final structures were not always consistent with those of a β -sheet, and this patch is not classified as such (see the discussion below). No dihedral angle- and H-bond constraints were applied to these residues during the calculation. We detected very few long-range NOEs involving residues 49–53 within the loop between $\beta 2$ and $\beta 3$. Unlike the other three loops flanked by the secondary structure elements, this loop showed a particularly poor degree of structural convergence (Figure 3a, 4d). Also, no long range NOEs were observed for the residues in the C-terminal tail.

Backbone Dynamics

That these two regions (loop 4 and the tail) lack long-range NOEs and did not converge to particular conformation is to be expected if they are highly mobile. In order to gain insight into the dynamics of this protein, we carried out backbone ¹⁵N relaxation measurements. Analysis of these data using the “model free” approach^{16,17} confirmed the general observations from our structural study (Figure 4). The order parameter (S^2) was mostly high (> 0.85) from the N-terminus up until residue 60. The residues for which we detected few or no long-range NOEs, such as residues 12–14, 49–53, and 61–81, have lower S^2 values. The correlation times for internal motions (τ_e) are also significantly higher for these residues (> 200 psec), together with the lower S^2 indicating higher degree of internal motions in these regions. These findings are mirrored in the variation in the averaged rmsd of backbone heavy atoms (C α , N, and CO) calculated from the 25 lowest-energy structures (Figure 4d).

The C-terminal 20 residues exhibit the highest mobility with S^2 below 0.2, τ_e in the 0.8 ns range, and mostly negative NOEs. We found that only in a few residues the chemical exchange term (R_{ex}) contributed significantly to the relaxation rate R_2 . Particularly, the high flexibility in the C-terminal region does not appear to result from such micro- to millisecond time scale

motions; R_{ex} values were uniformly small when the R_{ex} term is included in the fitting (data not shown).

PapI Is a Winged Helix-Turn-Helix Protein

A database search revealed that PapI is a member of the protein family classified as “FaeA-like” proteins (Pfam accession number 04703).^{10,11} This family includes several proteins that are known or hypothesized to participate in fimbriae regulation in concert with Lrp. While no structural study on this family of proteins has been reported, they are hypothesized to be winged helix-turn-helix (wHtH) proteins. This structural motif is a variation on the more common motif known as helix-turn-helix (HtH), which is made of $\alpha 1$ - $\alpha 2$ -turn- $\alpha 3$.^{18,19,20,21} The canonical fold of a winged HtH protein consists of an HtH (α - α -turn- α) followed by the wing (β -loop- β), and it is identical to the overall fold of PapI. A VAST search^{22,23} for its structural homologs confirmed that PapI is in fact a wHtH protein. Structure alignment of PapI with some of the known wHtH proteins (Table 2) showed the level of similarities expected only from closely related structures, based on the criteria previously described.²⁴

It is interesting to note that, in the three-stranded variety of winged HtH proteins, there is an additional short β -strand between the first two helices running antiparallel to the strand corresponding to $\beta 2$ in PapI.^{20,21} In PapI, these patches correspond to G15-K16 (in loop1) and Y56-W57 ($\beta 2$), for which we did detect an NOE pattern suggestive of an antiparallel β -sheet, as discussed above (broken arrows in Figure 3d). In some of our 25 final structures, the dihedral angles of this G15-K16 patch were such that it was predicted to be a β -strand. We did not classify it as β -strand since it was not predicted to be one in the majority of the cases. Their $C\alpha$ chemical shift index did not show consistent β -strand character (Figure 3c), and the inter-strand NOEs were generally weak. J-coupling constants for these residues were in the random-coil range (6.6 Hz for G15 and 6.5 Hz for K16). The dihedral-angle prediction by the TALOS program³² classified the middle residue of the G14-G15-K16 stretch as “unclassified,” and K16 in G15-K16-T17 as “good” but its predicted ϕ angle was $-86 \pm 18^\circ$, which is on the edge of the β -strand region. It is worth noting that the dynamics data for the residues immediately before G15-K16 suggest higher degree of mobility than average for this region, shown as lowered S^2 and raised τ_e values.

PapI Is a non-specific DNA-binding Protein

Most of HtH proteins and wHtH proteins studied to date are either known or proposed DNA-binding proteins.^{20,21} It was reported that PapI enhanced the affinity of Lrp to its substrate DNA in a sequence-specific manner. Of the six Lrp-binding sites in the *pap* promoter region, site 2 and site 5, both of which include the GATC dam site, showed pronounced PapI response, with site 5 showing a significantly stronger response than site 2. PapI did not appear to affect the binding of Lrp to the remaining four sites.⁹ Although previous attempts to detect the DNA-binding activity of PapI independent of Lrp, as assayed by electrophoresis mobility shift and UV-crosslinking, were inconclusive,⁹ knowledge of its structure prompted us to revisit this issue.

We tested whether PapI possesses its own DNA-binding capacity by titrating it with *pap* site 5 and site 2 DNA. Figure 5a shows the typical result of such experiments, in which 1H - ^{15}N -TROSY HSQC spectra of deuterated PapI were collected at different *pap* 5 DNA concentrations and overlaid with the DNA-free spectrum of PapI. We detected a number of residues affected by the presence of DNA, as evidenced by the changes in their amide chemical shifts. To confirm that these changes are indeed due to the interaction with DNA, DNase I was added to some of the samples. In all such cases, the original (DNA-free) spectra were restored in a matter of minutes (data not shown), leading us to believe that PapI is capable of binding DNA in the absence of Lrp.

In an attempt to estimate the dissociation constant (K_d) of this interaction, the amide ^1H or ^{15}N resonance displacements from the free positions of several representative residues were plotted as a function of DNA concentration (Figure 5b), and the binding curves were fitted with K_d and the bound state chemical shift (CS_b) as adjustable parameters. Interestingly, for all residues examined, the data fit poorly with a model assuming a single PapI molecule bound per DNA molecule. The model involving two molecules of PapI bound per DNA duplex significantly improved the fit. As shown in Figure 5b, the data at low DNA concentration suggest nearly stoichiometric (2-to-1) binding. At higher DNA concentrations, the binding seems somewhat weaker, due possibly to some cooperativity between PapI molecules at low DNA that is lost when the DNA level is higher relative to protein. It is difficult to estimate the K_d with great confidence without knowing the precise stoichiometry, but our preliminary analysis based on the concentration dependence of the observed chemical shift changes suggests a K_d in the 10 μM range (in the presence of 150 mM KCl and 10 mM MgCl_2).

When the same experiment was repeated with *pap 2* DNA, we saw little or no difference in its affinity. Moreover, PapI was bound to a randomized *pap* site 5 sequence (a negative control) with roughly the same affinity, and the residues responsible for the binding appeared to be the same in all three cases (data not shown). Under a high salt concentration shown in Figure 5a (150 mM KCl/10 mM MgCl_2), the movements and intensities of the majority of amide resonances throughout the titration clearly indicate that the binding was within the fast exchange regime. Since there was no measurable difference in K_d , we attempted to detect the difference in the exchange rates by lowering salt concentration. We did see an apparent decline in K_d due to the lower salt concentration but failed to detect any measurable differences in the exchange rates between *pap* 5 site, *pap* 2 site, and randomized sequence at a salt concentration as low as 20 mM KCl/0.5 mM MgCl_2 (data not shown). We also looked for qualitatively different changes in chemical shifts caused by different DNA pieces as a sign of sequence-specific binding, since different bases should affect PapI residues differently, even when the same residues were responsible for the interaction. We did not see, at least as assayed by ^1H - ^{15}N correlation spectra, significant differences between PapI spectra collected with different DNA sequences. From these observations, we propose PapI is a non-specific DNA-binding protein, at least in the absence of Lrp.

DNA-binding Sites of PapI

We attempted to locate the DNA-binding surface of PapI by comparing the magnitude of the amide chemical shift changes in PapI upon binding DNA (Figure 5c). We found three separate patches of residues most severely affected by DNA: (1) Q45, R46, S47, R50, R51, G52, M53, A54, T55, F58, (2) Y29, Q30, and (3) A18. Residues in the first category are within the tip of the wing, which corresponds to the residues 48 through 55, or nearby. Amide resonances of R46, G52, and T55 could not be traced beyond the very early stages of titration, due presumably to larger chemical shift changes and had to be extrapolated based on the available data points. The second category, Y29 and Q30 are on the outer surface of $\alpha 3$ and highly conserved among the FaeA-like proteins (Figure 7). A18 is far from the other two patches in the primary sequence but located on $\alpha 2$, positioned between $\alpha 3$ and the tip of the wing. When the chemical shift changes were color-coded and plotted in a ribbon diagram of PapI structure, it reveals an elongated cluster of residues forming a DNA binding surface (Figure 5d). This area also coincides with a unique surface with high positive charge density and electrostatic potential (Figure 5e). We did not see residues in the flexible C-terminal tail being significantly affected by DNA.

Discussion

Modulation in the DNA-binding affinity of Lrp by PapI plays the central role in the phase-variation mechanism that regulates the *papBA* operon.^{1,2} In phase-OFF cells, Lrp is bound to the proximal site cluster (sites 1, 2, and 3). Lrp is bound to the distal cluster (sites 4, 5, and 6) in phase-ON cells. It has been established that PapI makes direct contact with Lrp, and the PapI-Lrp complex binds significantly more tightly to *pap 5* and *pap 2* DNA than Lrp alone, while the binding of Lrp to other four *pap* sites did not seem to be enhanced by the presence of PapI. Also, methylation of DNA seems to affect specifically the ternary complex of DNA, PapI, and Lrp but not the binary complex of Lrp with DNA.⁹ These prior findings attest to the ability of PapI to either distinguish these subtle differences on its own or confer such ability to Lrp. It is also possible that the PapI-Lrp complex acquires additional features that do not exist in either one of the participants. The goal of this study was to understand, on the structural basis, the mechanism by which PapI accomplishes them.

DNA-binding mechanism

We found that PapI is a wHtH DNA-binding protein. The mechanisms by which HtH proteins bind to DNA have been elucidated in a number of previously reported cases. Typically, $\alpha 3$ (known as the “recognition helix”), placed against the major groove of DNA, forms the main contact surface, while other parts of the protein sometimes provide additional contacts. In winged HtH proteins, another major contact is provided by the tip of the “wing,” which is the loop between the β -strands that immediately follow the recognition helix.^{20,21} Also, it has been reported that, in cases where the winged HtH motif comprises the whole protein, the C-terminal tail can provide yet another contact surface.³³ In this study, the PapI residues most affected by the presence of DNA also converged mainly to the recognition helix and the wing (but not the C-terminal tail). It is tempting to think that PapI also binds to DNA through a similar mechanism, although we do not have experimental data to directly support this notion. Figure 6 shows a model for such a mechanism. In this figure, PapI is simply overlaid with the wHtH domain of *Staphylococcus aureus* MecI, one of the several wHtH proteins for which the DNA-bound structures are available (PDB ID: 1SAX).³⁴ As discussed above (Table 2), PapI shows a high degree of structural homology to MecI DNA-binding domain (rmsd 1.3Å for the structured residues), and the proposed DNA-binding surface of PapI, shown in Figure 6c, seems to be well positioned to make contacts with DNA.

We found that the structures of a number of wHtH proteins and PapI could be aligned with good agreement. While relative orientation of secondary structure elements is generally very good, however, there are minor variations within such elements themselves. For instance, the length of the wing is not consistent among wHtH proteins that are otherwise highly similar.²¹ The wings in most wHtH proteins consist of longer β -strands and a sharp turn between them, while these strands in PapI dissolve about half way through the wing and the tip of the wing actually appears to be a flexible loop, rather than a well-defined turn. The poor structural convergence of this region reflects the fact that we detected very few inter-residue NOEs. We believe this is a fair representation of its structure in solution, since the dynamics measurements confirmed that it was markedly more mobile than the flanking residues in the β -strand conformation (Figure 4). In the case of MecI, the wing is known to undergo a conformational change upon binding to DNA.^{31,34} We speculate the flexibility of the wing could facilitate the conformational changes necessary for DNA-binding. The length of the first two helices also seems to be somewhat variable. In the case of MecI-PapI alignment, the helix corresponding to PapI $\alpha 2$ is longer almost by a pitch in MecI, and the loops connecting $\alpha 2$ and $\alpha 3$ (the “turn”) in these proteins assume different conformations to maintain the relative orientation of these helices essentially identical.

The FaeA-like family

To date, twenty-one protein sequences have been identified as members of the FaeA-like family (Pfam accession number PF04703). These include fourteen PapI orthologs from various *E.coli* strains, five *Salmonella* orthologs, and one in archaeobacteria *Pyrobaculum aerophilum*. Of these, ten have been studied at the genetic or biochemical level and are known to regulate the expression of fimbriae adhesins in concert with Lrp orthologs.

Figure 7 shows the sequence alignment of PF04703. The wHtH motif is often seen in multi-domain proteins, fused with other functional domains or additional small motifs to facilitate oligomerization or interaction with DNA and other proteins. One of the interesting features of the PF04703 family is that it is a rather rare stand-alone wHtH motif. For instance, *E.coli* FaeA and the four members from *Salmonella*, if their fold is similar to PapI, do not extend far beyond the last β -strand of the wing. Interestingly, this tail-less *E.coli* ortholog functions in a manner seemingly identical to PapI with a different result; FaeA regulates K88 fimbriae, another surface adhesin, in an Lrp- and dam-dependent phase-variation mechanism not unlike the regulation of the *pap* operon. Unlike PapI-Lrp complex, however, FaeA-Lrp complex is a repressor of *fae* operon, and substitution of PapI with FaeA results in the activation of *fae* operon in vivo.^{35,36}

What is the function of the PapI C-terminal tail? In the case of eukaryotic histone H1, another stand-alone wHtH protein, its short C-terminal tail is thought to mediate protein-protein interactions.^{37,38} In hepatocyte nuclear factor 3 (HNF-3 γ), its short tail provides a DNA-contact surface.³³ We do not have any evidence that the PapI tail interacts with other parts of the protein, and it seems to have no apparent structural role. PapI appears to be a monomer in solution (our unpublished analytical centrifugation observation). We did not see clear signs of PapI oligomerization upon DNA-binding, and in any event, the tail certainly does not seem to be involved in it. We did not detect significant interaction between the tail and DNA. We are currently studying the PapI-Lrp interaction, and the tail does not appear to play a role (our unpublished data). To our knowledge, there is no other macromolecule thought to interact with PapI. One possibility that has not been explored is that the tail might be involved in integrating the external signal inputs into the phase variation. It is interesting that the expression of F165 (1) fimbriae, a *pap* homolog, is regulated by Lrp and FooI (100% sequence identity with PapI) and yet, unlike *pap* pili expression, is sensitive to leucine and alanine.^{39,40} This raises the possibility that there might be unidentified signaling events that affect the *pap* expression.

PapI-Lrp interaction and mechanism of *pap* regulation

As reported previously, the Lrp-PapI complex has different DNA-binding activity from that of Lrp by itself, and the effects of PapI is specific to *pap* site 5 and 2.^{7,9} Upon learning that PapI is a DNA-binding protein itself, we initially speculated that the PapI-Lrp complex binds to these two sites tighter than Lrp alone by virtue of having an additional DNA-binding motif specific to these sequences. Our findings that PapI bound to all DNA pieces tested with similar affinity, however, do not support this notion. We are puzzled by these results since the PapI-Lrp complex is known to bind to other four *pap* sites, 1,3,4, and 6, as tightly as Lrp by itself.⁹ Based on the lack of substrate specificity we observed in this study, we are inclined to think that PapI was in direct contact with DNA in such cases where PapI did not enhance DNA-binding of Lrp in prior studies. We suspect that this may not be a case of simple sum of the two independent DNA-binding activities but that PapI and/or Lrp, within the context of the complex, acquire the observed specificity that neither has on their own.

One explanation would be that there might be a certain arrangement of the two DNA-binding motifs that is recognized specifically by *pap* site 5 and 2 but not by other sites. A more complicated scenario would be that PapI-Lrp complex is not a rigid-body contact and that either

one or both DNA-binding proteins might be structurally altered in a way that confers the specificity and higher affinity to these two sites. Further experiments are needed to explore these and other possibilities, and we hope that PapI structure, chemical shift assignments, and new findings on PapI-DNA interaction presented here would contribute to such ends.

Materials and Methods

Expression and Purification of Wild-type and His-tagged PapI

All expression constructs used in this study were based on pET28b(+) (Novagen). To construct the his-tagged PapI-2, its coding sequence was PCR-amplified from pDAL283³ and cloned into pET28b(+) at its Nco I and Xho I sites in frame with the C-terminal his-tag and stop codon of the plasmid. These constructs include two residues coded by the Xho I restriction site, Leu74 and Glu75, between the C-terminal Thr73 and the his-tag. The second residue of the protein, which was mutated to a glutamate during the cloning, was mutated back to the wild-type lysine with site-directed mutagenesis. His-tagged PapI-1 and WT PapI-2 were constructed following the same procedure, except that the reverse primer for the WT PapI-2 contained the stop codon immediately following the C-terminal Thr73. Pfu Turbo DNA polymerase (STRATAGENE) was used for all PCR reactions. Restriction enzymes and T4 DNA ligase were obtained from New England BioLabs.

PapI was overexpressed in BL21 (DE3) (Novagen) by the addition of 1 mM IPTG during the mid-log growth phase. The culture was harvested upon saturation by centrifugation and disrupted by FrenchPress in 50 mM potassium phosphate at pH 7.4 containing 300 mM potassium chloride, 5 mM β -mercaptoethanol. His-tagged PapI was purified to apparent ~95% homogeneity with Ni-NTA column (QIAGEN) following the manufacturer's protocol. WT PapI was purified following the published protocol.³ Purified protein was dialyzed in 50 mM potassium phosphate pH 6.0, 150 mM potassium chloride, 5 mM β -mercaptoethanol, 100 μ M EDTA and concentrated to 400 μ M with Amicon Centriprep (Millipore). In our hands, this appeared to be the upper limit of PapI concentration under this particular condition. Protein concentration was determined by its UV absorbance at 280nm. PapI contains one tryptophan, four tyrosines, and two phenylalanines, and its extinction coefficient at 280nm is expected to be 10810 (cm²·M)⁻¹.⁴¹ Purified protein was stored at -80°C and dialyzed into the final buffers described under each experiment below.

Isotopical Labeling and NMR spectroscopy

Isotopically labeled proteins were grown in 1L M9 medium in pure H₂O or D₂O (Cambridge Isotope Laboratories) containing 1 mM magnesium sulfate, 0.1 mM calcium chloride, 0.02 mM iron(III) chloride, 0.5 μ g/ml thiamine, and 30 μ g/ml kanamycine, supplemented with 1 gram of ¹⁵N-ammonium chloride (Cambridge Isotope Laboratories) and the following amounts of glucose: 5 grams of unlabeled glucose for ¹⁵N- and ¹⁵N/²H labeled samples, 2 grams of ¹³C-labeled glucose (Cambridge Isotope Laboratories) for ¹⁵N/¹³C/¹H labeled samples. All protein samples were purified and concentrated as described above. D₂O and sodium azide were added to the NMR samples to 8% and 0.02%, respectively.

All NMR spectra were collected at 25°C on a Varian Inova 600 MHz spectrometer equipped with either a room-temperature or cryogenic ¹H/¹³C/¹⁵N triple resonance probe with a Z-direction pulsed field gradient and a separate ²H channel for locking or ²H decoupling. NMR data were analyzed with the nmrPipe package.⁴² Assignments were made with ANSIG3.3.^{43,44}

For the backbone- and sidechain assignments, the following spectra were collected using an ¹⁵N/¹³C-labeled sample: 2D ¹⁵N-HSQC,¹² 2D ¹³C-CT HSQC,⁴⁵ 3D HNCACB,⁴⁶ (H)C

(CO)NH-TOCSY, H(C)(CO)NH-TOCSY,^{47,48,49} and HNCO,^{12,46} and 4D ¹⁵N, ¹³C-edited NOESY.⁵⁰ Additionally, 4D ¹³C, ¹³C-edited NOESY⁵¹ was collected for the structure determination. The sample used for this experiment was prepared by exchanging the ¹⁵N/¹³C-labeled sample used for the assignments into a D₂O-based buffer (50 mM potassium phosphate pD 6.0, 150 mM KCl, 2 mM DTT) with a spin column containing 5ml Sephadex G25 resin (Sigma) equilibrated in this buffer. Also, 3D ¹⁵N-edited NOESY⁵² and 3D HNHA⁵³ were collected with an ¹⁵N-labeled sample. The mixing time for all NOESY spectra was set to 150 milliseconds.

A deuterated, ¹⁵N-labeled sample was used in DNA-binding experiments. Purified and concentrated protein was dialyzed with or without DNA in 50 mM potassium phosphate pH 6.0, 150 mM potassium chloride, 10 mM magnesium chloride, 5 mM β-mercaptoethanol. It appeared that all amide deuterons had been sufficiently exchanged with protons during the purification, as judged from resonance intensities from deuterated and protonated protein samples, and no further effort was made to accelerate the deuteron exchange process. DNA sequence and the concentration of protein and DNA are described under “DNA-binding experiments” below. The effect of DNA-binding on PapI was monitored by collecting ¹H-¹⁵N TROSY-HSQC spectra⁵⁴ throughout the course of the titration.

Structure Calculations

Structure calculations for PapI were performed with the program X-PLOR 3.851¹³ using experimental NOE constraints, experimental and theoretical dihedral angle constraints, and simulated H-bond constraints. Based on the integrated peak intensities, all NOEs were classified into one of the three classes, weak, medium, and strong, corresponding to the upper distance limit of 5.2 Å, 4.2 Å, and 3.2 Å, respectively. The lower distance limit was set to 1.8 Å for all classes. Ambiguous NOEs were applied to all sidechain protons that were not stereospecifically assigned. In total, 1153 NOE constraints were used, of which 443 were weak, 432 were medium, and 278 were strong.

Backbone dihedral angle constraints and H-bond constraints were applied only to the residues that are clearly engaged in helical or β-strand structures. These residues were chosen based on their Cα chemical shift index,¹⁴ NOE patterns indicative of α-helix or β-strand,¹⁵ and ³J_{HaHN} coupling constants.^{15,53} The coupling constants derived from the HNHA spectrum were converted into ϕ angles as described previously.⁵³ These calculated ϕ_i values were applied as constraints with an error margin of ±10°. Theoretical ψ angle constraints of -47° for α-helix and 135° for β-sheet were applied with ±30° margin of error. The dihedral angles predicted by the TALOS program³² were used only to confirm our results were reasonably consistent with them but not used as the constraints. Two distance constraints that simulate the H-bond geometry were applied for each pair of residues involved in an H-bond: 1.95±0.35 Å between a carbonyl oxygen and an amide proton, and 2.90±0.40 Å between a carbonyl oxygen and an amide nitrogen.

Structure calculations were started by applying the distance-geometry substructure-embedding protocol to an extended PapI peptide with ideal local geometry. Fifty structures were generated from this procedure using all experimental NOEs. These structures were then regularized using a distance-geometry simulated-annealing protocol, followed by repeated simulated-annealing refinement. Dihedral angle- and H-bond constraints were introduced in the final rounds of refinement. Twenty-five structures (of the 50 total) with the lowest energy and without significant constraint violations were selected for further analysis. The alignment and rmsd of final structures were calculated with MOLMOL 2K.2.⁵⁵ The average structure was calculated and subjected to Powell minimization with all constraints.

Backbone ^{15}N Relaxation

Longitudinal and transverse relaxation times T_1 and T_2 (or the inverse rates R_1 and R_2), and $\{^1\text{H}\}^{15}\text{N}$ NOE factors of backbone amide ^{15}N nuclei were measured using inverse-detected 2D experiments.^{56,57,58} The following relaxation delays were used: 11.0, 55.0, 132.0, 231.0, 374.0, 550.0, 748.0, and 935.0 msec for T_1 measurements, and 16.6, 33.2, 66.3, 82.9, 99.5, 132.6, 165.8, and 198.9 msec for T_2 measurements. The relaxation rates R_1 and R_2 were extracted by fitting the peak intensities with a single exponential decay function. The $\{^1\text{H}\}^{15}\text{N}$ NOE factor was taken as the ratio of the peak intensities with and without ^1H saturation during 5 sec of the total 8 sec recycle delay period.

The R_1 , R_2 and NOE data were analyzed with the “model free” approach,^{16,17} using the program ModelFree4.^{57,58} We assumed the overall tumbling of the molecule is approximately isotropic, governed by a single rotational correlation time τ_m . Fast internal motions on the psec time scale are described by the order parameter (S^2) and correlation time τ_e . The equations and parameters relating R_1 , R_2 , and NOE values with τ_m , S^2 and τ_e were identical to the equations 1–3 in Kay et al.⁵⁶ To obtain the overall correlation time, we fit individual T_1/T_2 ratios with τ_m as the single variable while ignoring contributions from fast internal motions. The average of individual τ_m values from residues in well-structured regions were taken as the global rotational correlation time. The order parameter (S^2) and correlation time for internal motions (τ_e) were extracted for each residue with τ_m fixed at the average value. Additionally, relaxation data were also analyzed with the addition of a term to R_2 describing exchange broadening (R_{ex}) on the μsec to msec time scale to assess contributions from such motions. R_{ex} terms were extracted along with S^2 and τ_e while τ_m was fixed.

DNA-binding Experiments

DNA-binding experiments were done with deuterated, ^{15}N -labeled his-tagged PapI-2. WT PapI was used only to confirm that the results were comparable. PapI was titrated with DNA pieces containing the Lrp-binding sites of the *pap* promoter region. The following DNA primers (and their complementary strands) were obtained from Independent DNA Technologies and annealed to form double-stranded DNA as described previously:⁹

pap site 5: 5'- TCATTTAGACGATCTTTTATGCTGTAAATTCA -3'

pap site 5 shuffled: 5'-ATAATGGTTTAACTTAATGTTCTTCGCCAATT-3'

pap site 2: 5'- TGATTTAAACGATCTTTTAACCCACAAAACAA -3'

For each DNA piece, two NMR samples were prepared: one with the saturating amount of DNA (650 μM) and the other without DNA, both of which contained the same amount of PapI (180 μM). Both samples were dialyzed simultaneously in the same buffer (50 mM potassium phosphate pH 6.0, 150 mM potassium chloride, 10 mM magnesium chloride, 5 mM β -mercaptoethanol), and the volume of the samples was adjusted to be equal prior to data collection. To achieve desired concentration of DNA in both samples, appropriate volume of samples were exchanged between the high- and low-DNA samples. At each protein/DNA ratio, a ^1H - ^{15}N TROSY spectrum⁵⁴ was collected and overlaid with the spectrum from the free protein to monitor the peak movements.

The chemical shift perturbation between the DNA-bound and free PapI was calculated to locate the DNA-binding surface. It was normalized by the following formula and expressed in the unit of Hz:

$$\text{Chemical shift perturbation} = \sqrt{(\Delta\delta^1\text{H} \times 600)^2 + (\Delta\delta^{15}\text{N} \times 60.68)^2},$$

where $\Delta\delta^1\text{H}$ and $\Delta\delta^{15}\text{N}$ are the differences in chemical shifts between the initial and final data points of the titration in proton and nitrogen, respectively.

DNA-binding Constant Calculation

The overall dissociation constant K_d for the PapI-DNA interaction was estimated by taking the average of K_d values calculated for several residues with the largest movements during the titration by the following procedure. For each of these residues, the chemical shift of either amide ^1H or ^{15}N was described as the weighted average shifts of the free and bound states under two-site, fast exchange condition:

$$\text{CS}(x) = \text{CS}_0(1 - f(x)) + \text{CS}_b f(x),$$

where CS_0 and CS_b are the chemical shifts in the free and bound states, respectively. $f(x)$ is the fraction bound at each total DNA concentration (x) and relates to K_d , total PapI and DNA concentrations by the following equation:

$$f(x) = \left(\frac{1}{2}\right) \left(\frac{K_d + x}{[\text{PapI}_{\text{total}}]} + 1 \right) - \left(\frac{1}{2}\right) \sqrt{\left(\frac{K_d + x}{[\text{PapI}_{\text{total}}]} + 1 \right)^2 - 4 \frac{x}{[\text{PapI}_{\text{total}}]}}.$$

The binding curve was fitted with the Gnuplot program with K_d and CS_b as adjustable parameters. The data were fitted either by a single binding site (on DNA) model or a model involving two binding sites (see text).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Hernday A, Braaten B, Low D. The intricate workings of a bacterial epigenetic switch. *Adv. Exp. Med. Biol* 2004;547:83–89. [PubMed: 15230094]
2. Hernday A, Krabbe M, Braaten B, Low D. Self-perpetuating epigenetic pili switches in bacteria. *Proc. Natl. Acad. Sci. USA* 2002;99:16470–16476. [PubMed: 12202745]
3. Nou X, Skinner B, Braaten B, Blyn L, Hirsch D, Low D. Regulation of pyelonephritis-associated pili phase-variation in *Escherichia coli*: binding of the PapI and the Lrp regulatory proteins is controlled by DNA methylation. *Mol. Microbiol* 1993;7:545–553. [PubMed: 8096319]
4. Braaten BA, Nou X, Kaltenbach LS, Low DA. Methylation patterns in pap regulatory DNA control pyelonephritis-associated pili phase variation in *E. coli*. *Cell* 1994;76:577–588. [PubMed: 7906204]
5. Braaten BA, Blyn LB, Skinner BS, Low DA. Evidence for a methylation-blocking factor (mbf) locus involved in pap pilus expression and phase variation in *Escherichia coli*. *J. Bacteriol* 1991;173:1789–1800. [PubMed: 1671857]
6. Nou X, Braaten B, Kaltenbach L, Low DA. Differential binding of Lrp to two sets of pap DNA binding sites mediated by Pap I regulates Pap phase variation in *Escherichia coli*. *EMBO J* 1995;14:5785–5797. [PubMed: 8846772]
7. Kaltenbach LS, Braaten BA, Low DA. Specific binding of PapI to Lrp-pap DNA complexes. *J. Bacteriol* 1995;177:6449–6455. [PubMed: 7592419]
8. Baga M, Goransson M, Normark S, Uhlin BE. Transcriptional activation of a pap pilus virulence operon from uropathogenic *Escherichia coli*. *EMBO J* 1985;4(13B):3887–3893. [PubMed: 2868893]
9. Hernday AD, Braaten BA, Low DA. The mechanism by which DNA adenine methylase and PapI activate the pap epigenetic switch. *Mol. Cell* 2004;12:947–957. [PubMed: 14580345]
10. Sonnhammer EL, Eddy SR, Birney E, Durbin R. Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Res* 1998;26:320–322. [PubMed: 9399864]
11. Sonnhammer EL, Eddy SR, Durbin R. Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* 1998;28:405–420. [PubMed: 9223186]

12. Kay LE, Keifer EP, Saarinen T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc* 1992;114:10663–10665.
13. Brünger, AT. A System for X-ray Crystallography and NMR. New Heaven, CT: Yale University Press; 1992. X-PLOR, version 3.1.
14. Wishart DS, Sykes BD. The ^{13}C Chemical-Shift Index: A simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR* 1994;4:171–180.
15. Wüthrich, K. NMR of Proteins and Nucleic Acids. chapter 9. New York, NY: Wiley Interscience; 1986.
16. Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc* 1982;104:4546–4559.
17. Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc* 1982;104:4559–4570.
18. Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM. The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *FEMS Microbiol. Rev* 2005;29:231–262. [PubMed: 15808743]
19. Huffman JL, Brennan RG. Prokaryotic transcription regulators: more than just the helix-turn-helix motif. *Curr. Opin. Struct. Biol* 2002;12:98–106. [PubMed: 11839496]
20. Gajiwala KS, Burley SK. Winged helix proteins. *Curr. Opin. Struct. Biol* 2000;10:110–116. [PubMed: 10679470]
21. Brennan RG. The winged-helix DNA-binding motif: another helix-turn-helix takeoff. *Cell* 1993;74:773–776. [PubMed: 8374950]
22. Gibrat JF, Madej T, Bryant SH. Surprising similarities in structure comparison. *Curr. Opin. Struct. Biol* 1996;6:377–385. [PubMed: 8804824]
23. Madej T, Gibrat JF, Bryant SH. Threading a database of protein cores. *Proteins* 1995;23:356–369. [PubMed: 8710828]
24. Flores TP, Orengo CA, Moss DS, Thornton JM. Comparison of conformational characteristics in structurally similar protein pairs. *Protein Sci* 1993;2:1811–1826. [PubMed: 8268794]
25. Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Marchler GH, Mullokandov M, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Yamashita RA, Yin JJ, Bryant SH. CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res* 2005;33Database Issue:D192-6
26. Ye J, Kandegedara A, Martin P, Rosen BP. Crystal structure of the *Staphylococcus aureus* p1258 CadC Cd(II)/Pb(II)/Zn(II)-responsive repressor. *J. Bacteriol* 2005;187:4214–4221. [PubMed: 15937183]
27. Eicken C, Pennella MA, Chen X, Koshlap KM, VanZile ML, Sacchettini JC, Giedroc DP. A metal-ligand-mediated intersubunit allosteric switch in related SmtB/ArsR zinc sensor proteins. *J. Mol. Biol* 2003;333:683–695. [PubMed: 14568530]
28. Ray SS, Bonanno JB, Chen H, de Lencastre H, Wu S, Tomasz A, Burley SK. X-ray structure of an *M. jannaschii* DNA-binding protein: implications for antibiotic resistance in *S. aureus*. *Proteins* 2003;50:170–173. [PubMed: 12471609]
29. Schade M, Turner CJ, Kuhne R, Schmieder P, Lowenhaupt K, Herbert A, Rich A, Oschkinat H. The solution structure of the Zalpha domain of the human RNA editing enzyme ADAR1 reveals a prepositioned binding surface for Z-DNA. *Proc. Natl. Acad. Sci. USA* 1999;96:12465–12470. [PubMed: 10535945]
30. Pohl E, Haller JC, Mijovilovich A, Meyer-Klaucke W, Garman E, Vasil ML. Architecture of a protein central to iron homeostasis: crystal structure and spectroscopic analysis of the ferric uptake regulator. *Mol. Microbiol* 2003;47:903–915. [PubMed: 12581348]
31. Garcia-Castellanos R, Marrero A, Mallorqui-Fernandez G, Potempa J, Coll M, Gomis-Ruth FX. Three-dimensional structure of MecI. Molecular basis for transcriptional regulation of staphylococcal methicillin resistance. *J. Biol. Chem* 2003;278:39897–39905. [PubMed: 12881514]
32. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* 1999;13:289–302. [PubMed: 10212987]

33. Clark KL, Halay ED, Lai E, Burley SK. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* 1993;364:412–420. [PubMed: 8332212]
34. Garcia-Castellanos R, Mallorqui-Fernandez G, Marrero A, Potempa J, Coll M, Gomis-Ruth FX. On the transcriptional regulation of methicillin resistance: MecI repressor in complex with its operator. *J. Biol. Chem* 2004;279:17888–17896. [PubMed: 14960592]
35. Huisman TT, Bakker D, Klaasen P, de Graaf FK. Leucine-responsive regulatory protein, IS1 insertions, and the negative regulator FaeA control the expression of the fae (K88) operon in *Escherichia coli*. *Mol. Microbiol* 1994;11:525–536. [PubMed: 8152376]
36. Huisman TT, de Graaf FK. Negative control of fae (K88) expression by the 'global' regulator Lrp is modulated by the 'local' regulator FaeA and affected by DNA methylation. *Mol. Microbiol* 1995;16:943–953. [PubMed: 7476191]
37. Cerf C, Lippens G, Ramakrishnan V, Muyldermans S, Segers A, Wyns L, Wodak SJ, Hallenga K. Homo- and heteronuclear two-dimensional NMR studies of the globular domain of histone H1: full assignment, tertiary structure, and comparison with the globular domain of histone H5. *Biochemistry* 1994;33:11079–11086. [PubMed: 7727360]
38. Staynov DZ, Crane-Robinson C. Footprinting of linker histones H5 and H1 on the nucleosome. *EMBO J* 1988;7:3685–3691. [PubMed: 3208745]
39. Berthiaume F, Crost C, Labrie V, Martin C, Newman EB, Harel J. Influence of L-leucine and L-alanine on Lrp regulation of foo, coding for F1651, a Pap homologue. *J. Bacteriol* 2004;186:8537–8541. [PubMed: 15576806]
40. Crost C, Harel J, Berthiaume F, Garrivier A, Tessier MC, Rakotoarivonina H, Martin C. Influence of environmental cues on transcriptional regulation of foo and clp coding for F165(1) and CS31A adhesins in *Escherichia coli*. *Res. Microbiol* 2004;155:475–482. [PubMed: 15249065]
41. Gill SC, von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem* 1989;182:319–326. [PubMed: 2610349]
42. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 1995;6:277–293. [PubMed: 8520220]
43. Kralis PJ. ANSIG: A Program for the Assignment of Protein 1H 2D NMR Spectra by Interactive Graphics. *J. Magn. Reson* 1989;84:627–633.
44. Kralis PJ, Domaille PJ, Campbell-Burk SL, van Aken T, Laue ED. Solution Structure and Dynamics of Ras p21.GDP Determined by Heteronuclear Three- and Four-Dimensional NMR Spectroscopy. *Biochemistry* 1994;33:3515–3531. [PubMed: 8142349]
45. Vuister GW, Bax A. Resolution enhancement and spectral editing of uniformly ¹³C enriched proteins by homonuclear broadband ¹³C-¹³C decoupling. *J. Magn. Reson* 1992;98:428–435.
46. Muhandiram DR, Kay LE. Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson. Ser. B* 1994;103:203–216.
47. Montelione GT, Lyons BA, Emerson SD, Tashiro M. An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. *J. Am. Chem. Soc* 1992;114:10974–10975.
48. Grzesiek S, Anglister J, Bax A. Correlation of backbone amide and aliphatic side-chain resonances in ¹³C/¹⁵N-enriched proteins by isotopic mixing of ¹³C magnetization. *J. Magn. Reson. B* 1993;101:114–119.
49. Logan TM, Olejniczak ET, Xu RX, Fesik SW. A general method for assigning NMR spectra of denatured proteins using 3D HC(CO)NH-TOCSY triple resonance experiments. *J. Biomol. NMR* 1993;3:225–231. [PubMed: 8477187]
50. Muhandiram DR, Guang YX, Kay LE. An enhanced-sensitivity pure absorption gradient 4D ¹⁵N, ¹³C-edited NOESY experiment. *J. Biomol. NMR* 1993;3:463–470.
51. Vuister GW, Clore GM, Gronenborn AM, Powers R, Garrett DS, Tschudin R, Bax A. Increased Resolution and Improved Spectral Quality in Four-Dimensional ¹³C/¹³C-Separated HMQC-NOESY-HMQC Spectra Using Pulsed Field Gradients. *J. Magn. Reson. B* 1993;101(2):210–213.
52. Zhang O, Kay LE, Olivier JP, Forman-Kay JD. Backbone ¹H and ¹⁵N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J. Biomol. NMR* 1994;4:845–858. [PubMed: 7812156]

53. Vuister GW, Bax A. Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in ¹⁵N-enriched proteins. *J. Am. Chem. Soc* 1993;115:7772–7777.
54. Pervushin K, Riek R, Wider G, Wüthrich K. Attenuated *T*₂ relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. USA* 1997;94:12366–12371. [PubMed: 9356455]
55. Koradi R, Billeter M, Wuthrich K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph* 1996;14:51–55. [PubMed: 8744573]
56. Kay LE, Torchia DA, Bax A. Backbone dynamics of proteins as studied by ¹⁵N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* 1989;28:8972–8979. [PubMed: 2690953]
57. Palmer AG 3rd, Rance M, Wright PE. Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterized by proton-detected natural abundance carbon-13 heteronuclear NMR spectroscopy. *J. Am. Chem. Soc* 1991;113:4371–4380.
58. Mandel AM, Akke M, Palmer AG 3rd. Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol* 1995;246:144–163. [PubMed: 7531772]

Figure 1.

A simplified view of the *pap* promoter region. The cooperative binding of Lrp to the proximal sites inhibits transcription of *papBA* genes, and GATC^{dist} is methylated in phase-OFF cells. In phase-ON cells, PapI interacts with Lrp and promotes its binding to the distal sites through an unknown mechanism. GATC^{prox} is open for methylation, and transcription can be turned on.

Figure 2.

¹H-¹⁵N HSQC spectrum of PapI labeled with the backbone assignments. The spectrum was collected at 25°C with a ¹⁵N-¹³C-labeled sample in 50 mM potassium phosphate pH 6.0, 150 mM KCl, 5 mM β-mercaptoethanol, 100 μM EDTA, 8% D₂O, 0.02% sodium azide.

Assignment was made from Asn3 through His77 (except Pro48). Asterisks indicate residues aliased in the ^{15}N dimension. Boxed peaks are unassigned sidechains. Trp57 HN ϵ 1, located at 10.1 ppm along the ^1H dimension and 112.1 ppm along the ^{15}N dimension, is not shown.

Figure 3.

Results of the structure calculation. (a) Superposition of the 25 lowest-energy structures. Their backbone rmsd was $0.40\text{\AA}$ over the structured parts of the protein (residues 2–11, 17–24, 28–40, 44–47, and 56–59). This figure shows only the residues 1 through 73 (the C-terminal residue in the WT protein), and the 8 extra residues at the C-terminus of this construct are not shown. (b) The ribbon diagram of the averaged-minimized PapI structure (PDB ID: 2HTJ). Its unstructured C-terminal tail (residues 63–81) is not shown. (c) The observed C_{α} chemical shifts of residues 2–77 were compared with their respective random-coil chemical shifts, and their differences were plotted. With respect to the random-coil values, consecutive downfield shifts (positive trends) indicate α -helices, and upfield shifts indicate β -strands (negative trends). (d) An NOE pattern indicative of antiparallel β -sheets was observed between V44–S47 and Y56–L59 patches. A similar but weaker interaction was detected between Y56–W57 and G15–K16 (broken arrows).

Figure 4.

Backbone ^{15}N relaxation data. (a) The order parameters (S^2), (b) effective correlation time (τ_e), and (c) $\{^1\text{H}\}^{15}\text{N}$ -NOE are shown, along with (d) the averaged rmsd calculated from the backbone non-hydrogen atoms of the 25 lowest-energy structures shown in Figure 3a. In (d), rmsd for residues 78–81, for which dynamics data are unavailable, is not shown.

Figure 5.

The DNA-binding experiments. A deuterated, ^{15}N -labeled PapI sample was titrated with the *pap 2*, *pap 5*, and randomized *pap 5* duplex DNA. (a) A part of the ^1H - ^{15}N correlation spectra of PapI corrected at different *pap 5* DNA concentrations are overlaid. Only the 0 μM -, 35 μM -, 90 μM -, and 660 μM -DNA data points are shown for clarity. T17 resonance was broadened to the extent that it cannot be seen at this contour level. (b) Titration curves were constructed for several representative residues, including Q45 ^1H shown here. A model assuming a 1-to-1 stoichiometry did not fit the data (red line). The model involving 2-PapI-to-1-DNA improved the fit (green line). This was a consistent trend shared by all residues examined. (c) The chemical shift difference between the DNA-free state and saturated state was plotted across the PapI sequence. Data for P48 and T55 are not available. (d) The total chemical shift perturbation in the presence of DNA shown in (c) was color-coded and plotted onto the PapI structure. Red color indicates large chemical shift differences. Residues in blue color had smaller shift changes. (e) The electrostatic potential map of PapI. Basic residues are colored blue, and acidic residues are colored red. In (d) and (e), only residues 1–62 are shown.

Figure 6.
A model for PapI-DNA interaction. (a) PapI residues 1–62 are superimposed with the DNA-binding domain of *Staphylococcus aureus* MecI complexed with the *mec* promoter (PDB ID: 1SAX) to 1.3Å rmsd. (b) MecI dimer was removed and the figure is slightly rotated relative to (a) for clarity.

Figure 7. Sequence alignment of the FaeA-like family (PF04703). Black-colored residues are highly conserved. Gray-colored positions are occupied by functionally similar residues. The PapI secondary structure (H = α -helix, S = β -strand) is shown underneath. Abbreviations: *S.* = *Salmonella*, *E.* = *Escherichia*, *P.* = *Pyrobaculum*, hyp. = hypothetical.

Table 1

Statistical evaluation of the structure calculation

Rmsd from the distance constraints () and dihedral angle constraints (...)	final 25 structures	2HTJ ^a
long range NOE $ i-j >4$ (226)	0.0133 - 0.0033	0.0070
medium range NOE $1\leq i-j \leq 4$ (466)	0.0303 - 0.0023	0.0260
intraresidue NOE (461)	0.0121 - 0.0030	0.0070
hydrogen bond (22)	0.0130 - 0.0033	0.0100
dihedral ϕ (36) ψ (36)	0.50 - 0.08	0.2923
Rmsd from the ideal geometry used in X-PLOR		
bond ()	0.0021 - 0.0001	0.0017
angle (...)	0.49 - 0.010	0.4710
improper angle (...)	0.37 - 0.010	0.3470
Coordinate rmsd () for the non-hydrogen atoms in the final 25 structures	backbone	sidechain
vs the averaged structure	0.40 - 0.077	1.047 - 0.12
vs the averaged and minimized structure (2HTJ)	0.45 - 0.075	1.150 - 0.14

Numbers in parentheses indicate the total numbers of constraints in the categories.

^aThe averaged and minimized structure derived from the final 25 structures.

Table 2
Structural neighbors of PapI (selected entries from the VAST search results)

pdb chain	residues aligned	VAST score	rmsd (°) ^a	description	wHh domain ^b	reference
1U2W b	57 [39]	7.8 [7.8]	1.6 [1.3]	<i>S.aureus</i> CadC	ArsR	26
1RIU c	59 [39]	7.8 [8.0]	1.5 [1.1]	<i>S.aureus</i> CztA	ArsR	27
1KU9 a	58 [39]	7.4 [8.0]	1.7 [1.3]	<i>M.jannaschii</i> Mj1563	ArsR	28
1QGP a	56 [39]	7.1 [7.1]	1.0 [0.9]	Human Adar1	Z-alpha	29
1SFX a	57 [39]	6.8 [7.5]	1.5 [1.1]	<i>A.fulgidus</i> AF2008	ArsR	NA
1MZH a	59 [39]	6.7 [6.8]	1.6 [1.0]	<i>P.aeruginosa</i> Fur	FUR	30
2FBH a	53 [39]	7.3 [7.6]	1.5 [1.3]	<i>P.aeruginosa</i> Pa3341	MarR	NA
1OKR b	52 [36]	6.8 [7.3]	1.7 [1.3]	<i>S.aureus</i> Mec1	MarR	31

Search for the structural neighbors of PapI was done using its first 62 residues or only the 39 residues in the helices and strands (results in the square bracket).

^acalculated between the averaged-minimized PapI and target structures.

^bclassification by the NCBI Conserved Domain Database.²⁵

Abbreviations: S. = Staphylococcus, M. = Methanococcus, A. = Archaeoglobus, P. = Pseudomonas, NA: = not available