The structure of the yrdC gene product from *Escherichia coli* reveals a new fold and suggests a role in RNA binding

MARIANNA TEPOLOVA, VALENTINA TERESHKO, RUSLAN SANISHVILI, ANDRZEJ JOACHIMIAK, TATYANA BUSHUEVA, WAYNE F. ANDERSON, AND MARTIN EGLI

1Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee 37235
2Structural Biology Center, Biosciences Division, Argonne National Laboratory, Argonne, Illinois 60439
3Cardiology Research Center, Russian Academy of Medical Sciences, Moscow 121552, Russia
4Department of Molecular Pharmacology and Biological Chemistry, Northwestern University, Medical School, Chicago, Illinois 60611

(Received September 6, 2000; Final Revision September 29, 2000; Accepted October 9, 2000)

Abstract

The yrdC family of genes codes for proteins that occur both independently and as a domain in proteins that have been implicated in regulation. An example for the latter case is the *sua5* gene from yeast. *Sua5* was identified as a suppressor of a translation initiation defect in cytochrome *c* and is required for normal growth in yeast (Na JG, Pinto I, Hampsey M, 1992, Genetics 111:791–801). However, the function of the *Sua5* protein remains unknown; *Sua5* could act either at the transcriptional or the posttranscriptional levels to compensate for an aberrant translation start codon in the *cyc* gene. To potentially learn more about the function of *YrdC* and proteins featuring this domain, the crystal structure of the *YrdC* gene product from *Escherichia coli* was determined at a resolution of 2.0 Å. *YrdC* adopts a new fold with no obvious similarity to those of other proteins with known three-dimensional (3D) structure. The protein features a large concave surface on one side that exhibits a positive electrostatic potential. The dimensions of this depression, its curvature, and the fact that conserved basic amino acids are located at its floor suggest that *YrdC* may be a nucleic acid binding protein. An investigation of *YrdC*’s binding affinities for single- and double-stranded RNA and DNA fragments as well as tRNAs demonstrates that *YrdC* binds preferentially to double-stranded RNA. Our work provides evidence that 3D structures of functionally uncharacterized gene products with unique sequences can yield novel folds and functional insights.

Keywords: protein folding; RNA; structural genomics; *Sua5*; X-ray crystallography; *YrdC*

The *yrdC* gene from *Escherichia coli* and its homologs represent genes with unique sequence that were classified as a so-called cluster of orthologous groups of proteins (COG) (Tatusov et al., 2000). In *E. coli*, *yrdC* is present as an isolated entity and shows high similarity to the *yciO* gene from the same organism. Two families of genes contain portions that are homologous to *yrdC*; these are the *sua5* (the name refers to yeast) and the *hypF* (the name refers to *E. coli*) gene families (Fig. 1). Thus, *YrdC* constitutes a domain of *Sua5* from yeast, but *E. coli* lacks an ortholog of *Sua5* with a conserved domain architecture.

The *sua5* gene was identified as a suppressor of a translation initiation defect in the leader region of the iso-1–cytochrome *c* (*cyc1*) gene (Na et al., 1992). As a consequence of a single base-pair substitution, an aberrant ATG start codon is generated upstream and out of frame with the normal *cyc1* translation start codon (sua, suppressor of upstream ATG). *Sua5* enhances the amount of produced iso-1–cytochrome *c* from 2 to 60% of normal and the strain exhibits a slow growth phenotype, demonstrating that this gene is required for normal growth.

*Sua*-type suppressors of translation initiation defects can be divided into two classes (Hampsey et al., 1991; Pinto et al., 1992). The *sua7* and *sua8* suppressors shift the transcription start sites downstream of normal and therefore act at the transcriptional level. Conversely, *sua1* to *sua6* do not affect transcription start site selection and appear to use distinctly different mechanisms for restoring growth (Na et al., 1992). However, the gene does not display any unusual sequence motifs that would provide clues about the function of the protein it encodes. Thus, *Sua5* could alter the rate of *cyc1* transcription or, alternatively, it could act at the post-transcriptional level by affecting *cyc1* mRNA stability, AUG codon recognition, or ribosomal reinitiation (Na et al., 1992). In either case, *Sua5* would serve a regulatory role and may interact with DNA or RNA. Although no mammalian genes with sequences

Reprint requests to: Martin Egli, Department of Biological Sciences, Vanderbilt University, Box 1634, Station B, Nashville, Tennessee 37235; e-mail: martin.egli@vanderbilt.edu.
HypF protein is involved in H₂ stimulation of hydrogenase expression of hydrogenase production in metallocenter assembly. These results are complemented by studies featuring a YrdC domain, HypF (Fig. 1). In E. coli, the HypF protein is involved in H₂ stimulation of hydrogenase expression (Colbeau et al., 1993) and in the maturation of all three hydrogenases (Maier et al., 1996). Further experiments indicate that the HypF protein plays a role in that maturation of the HupSL and HupUV proteins in the photosynthetic bacterium Rhodobacter capsulatus and that the latter are involved in the cellular response to H₂ (Colbeau et al., 1998). Maturation of the large subunit of E. coli hydrogenase 3 is dependent on seven accessory proteins (Drapal & Bock, 1998). Although the YciA protease catalyzes C-terminal proteolytic cleavage of the large subunit, the HypA, HypB, HypC, HypD, HypE, and HypF proteins are required for metallocenter assembly. These results are complemented by studies of hydrogenase production in Bradyrhizobium japonicum (Olson & Maier, 1997) in which the hypF gene encodes a 750 amino acid protein that contains the two zinc finger motifs characteristic of the YrdF protein family (Fig. 1). A further characteristic of this family is a putative acylphosphatase domain at the N-terminus (Wright et al., 1998). As with the Sua5 protein above, however, the function of the HypF protein family within the determined regulatory pathways remains to be worked out.

Three-dimensional (3D) structures of proteins may furnish insight into function in cases where sequence information and genetic screening have not allowed assignment of a function to a gene product (Kim, 1998; Orengo et al., 1999; Teichmann et al., 1999). Several examples of structure determinations of proteins of unknown function were published in recent years (Colovos et al., 1998; Yang et al., 1998; Zarembinski et al., 1998; Cort et al., 1999; Hwang et al., 1999; Sinha et al., 1999; Voiz, 1999; Minasov et al., 2000). The small number of such studies completed to date does not permit one to draw general conclusions at this time. However, the work has already demonstrated the power of 3D structures to reveal evolutionary relationships that were not previously apparent at the sequence level (Hwang et al., 1999; Minasov et al., 2000). Moreover, structure analysis generally allows determination of enzyme active sites and, together with sequence alignments of homologous genes, identification of key residues involved in the binding of ligands or substrates. In one of the analyzed structures, a bound ATP hinted at a potential function of the protein (Zarembinski et al., 1998), and in a second one, biochemical assays guided by the structural findings showed the protein to have NTPase activity (Hwang et al., 1999). The important lesson from these structural genomics pilot studies is that in the majority of cases, the structure determinations have led to testable hypotheses with regard to the function of the investigated proteins.

To examine whether the structure of the YrdC gene product would shed light on its function, we determined the crystal structure of the YrdC protein from E. coli at high resolution. Here we discuss the potential benefits of structure analysis of a conserved protein domain of unique sequence with respect to discovery of new folding motifs and ideas that help resolve functional aspects.

**Results and discussion**

**Overview of the structure**

The crystal structure of the E. coli YrdC protein has been determined by the multi-wavelength anomalous diffraction technique (MAD) (Table 1) and refined to an R-factor of 20.2% at 2.0 Å resolution (Table 2). The structure reveals two independent YrdC molecules per crystallographic asymmetric unit and the final model comprises 186 (residues 3–188) and 184 (residues 4–187) amino acids for subunits 1 and 2, respectively. In addition to the six

Table 1. MAD data collection parameters (maximum resolution 1.96 Å, last shell 2.03–1.96 Å)

<table>
<thead>
<tr>
<th>Wavelength (Å)</th>
<th>Number of reflections</th>
<th>Completeness (%)</th>
<th>R_{sym} (%)</th>
<th>Phasing power λ_{s}/λ_{i} (Friedel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>λ₁ = 0.9795 (edge)</td>
<td>Measured</td>
<td>Unique</td>
<td>All (last shell)</td>
<td>6.0 (42.3)</td>
</tr>
<tr>
<td>λ₂ = 0.9794 (peak)</td>
<td>231,530</td>
<td>62,907</td>
<td>96.2 (99.5)</td>
<td></td>
</tr>
<tr>
<td>λ₃ = 0.9465 (reference)</td>
<td>230,966</td>
<td>62,780</td>
<td>94.7 (99.8)</td>
<td>6.2 (48.6)</td>
</tr>
<tr>
<td>Overall figure of merit: 0.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*R_{sym} = \frac{\sum_i \sum_h |I(hkl)| - \langle|I(hkl)|\rangle}{\sum_i \sum_h |I(hkl)|}\) for the intensity I of i observations of a reflection hkl.
C-terminal amino acids (8 in the case of subunit 2), the His tag is invisible. An example of the \( 2F_o - F_e \) sum electron density surrounding the final model is depicted in Figure 2. The two YrdC molecules adopt very similar conformations and the root-mean-square deviation (RMSD) between \( \alpha \)-carbon atoms in their backbones is 0.60 Å. The greatest conformational variation is found at their C-termini and involves residues G170–F187 \( \text{RMSD} \approx 1.39 \) Å; Fig. 3. In solution, the YrdC protein is in a monomeric state, as demonstrated by dynamic light scattering (data not shown).

The YrdC protein folds into a compact structure with overall dimensions 50 \times 30 \times 30 Å. A stereo diagram of the \( \alpha \)-carbon trace of subunit one is depicted in Figure 4, and secondary structure schematics for YrdC are shown in Figure 5. The fold can be classified as an \( \alpha \)0\( \beta \) twisted open-sheet structure with parallel and antiparallel orientations of adjacent \( \beta \)-strands. The structure comprises seven \( \alpha \)-helices and seven \( \beta \)-strands with an unusual order and topology of \( \beta \)-strands and very strong twisting between them. An \( \alpha \)-helical region at the N-terminus of the protein is followed by two antiparallel \( \beta \)-strands (\( \beta_1 \) and \( \beta_2 \)). These lead over to the second helix \( \alpha_2 \) that is then linked to \( \beta_3 \) by a five-residue loop. The third \( \beta \)-strand is located near the center of the protein, and a short turn leads over to \( \alpha_3 \) and \( \alpha_4 \) that form the other end of the protein together with \( \beta_4 \). An irregular region interrupted only by the single-turn helix \( \alpha_5 \) provides a connection to \( \beta_5 \), itself oriented parallel to \( \beta_3 \) and antiparallel to \( \beta_4 \). The loops at the carboxy ends of \( \beta_3 \) and \( \beta_5 \) connect them to \( \alpha \)-helices on opposite sides of the sheet (\( \alpha_3 \) and \( \alpha_6 \), respectively; Fig. 5).

Thus, \( \alpha_6 \) is inserted between \( \alpha_3 \) and \( \alpha_1 \) and leads back to the center of the molecule where \( \beta_6 \) is inserted antiparallel to \( \beta_3 \). A 12-residue loop links \( \beta_6 \) to \( \alpha_7 \) and a long loop interrupted by the short \( \beta_7 \) leads to the C-terminus. Considerable conformational flexibility of the C-terminal region is indicated by high temperature factors of residues there (46.0 Å² on average for G170 to R188, subunit 1; 42.6 Å² on average for G170 to F187, subunit 2) as well as by the presence of a glycine hinge (G170 and G171). This suggests that this portion of the YrdC protein may constitute part of an interdomain linker. It is

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>25.0–2.0 (bulk solvent correction)</td>
</tr>
<tr>
<td>Reflections ( (F \approx 2\sigma[F]) )</td>
<td>51,216</td>
</tr>
<tr>
<td>( R )-factor (%)</td>
<td>20.2</td>
</tr>
<tr>
<td>( R )-free (%)</td>
<td>21.3</td>
</tr>
<tr>
<td>Number of atoms</td>
<td>3,244 (2 YrdC, 2 phosphates, 386 H2O)</td>
</tr>
<tr>
<td>RMSDs</td>
<td></td>
</tr>
<tr>
<td>Bond length (Å)</td>
<td>0.009</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.5</td>
</tr>
<tr>
<td>Average temperature factors</td>
<td></td>
</tr>
<tr>
<td>Protein atoms (Å²)</td>
<td>25.4 (subunit 1), 23.8 (subunit 2)</td>
</tr>
<tr>
<td>Solvent molecules (Å²)</td>
<td>34.2</td>
</tr>
<tr>
<td>Phosphate ion (Å²)</td>
<td>49.5</td>
</tr>
</tbody>
</table>

*RMSDs from ideal geometry.

![Fig. 2. Quality of the final model of the *E. coli* YrdC crystal structure. Stereo diagram of the \( 2F_o - F_e \) Fourier sum electron density around the phosphate ion bound to K69, D74, and R82 in one of the YrdC subunits, drawn with TURBO FRODO (Cambillau & Roussel, 1997). The map is contoured at the 1σ level, atoms are colored yellow, red, blue, and orange for carbon, oxygen, nitrogen, and phosphorus, respectively, and oxygen atoms of water molecules are drawn as small red spheres.](image1)

![Fig. 3. Conformational variations between the two independent YrdC molecules per crystallographic asymmetric unit, drawn with the program MOLMOL (Koradi et al., 1996). The thickness of the line is proportional to the distance between corresponding \( \alpha \)-carbon atoms of individual molecules. The largest deviations are found between residues R172 and P175.](image2)
noteworthy that the conformations adopted by the C-termini of the YrdC molecules are not affected by crystal packing interactions in our structure. Thus, amino acids G170 to R188 in the case of molecule 2 at the C-terminus display distances to residues from adjacent protein molecules that exceed 10 Å.

Folding topology

The structure of the YrdC protein can be dissected into three parts. The first is composed of $\alpha_1$, $\beta_1$, $\beta_2$, and $\alpha_2$, and the second comprises $\beta_3$, $\alpha_3$, $\alpha_4$, $\beta_4$, $\alpha_5$, and $\beta_5$. These two portions are bridged by $\alpha_6$ and $\beta_6$ that constitutes the third part. The adjacent flexible C-terminal region including $\alpha_7$ and $\beta_7$ is folded down over all three parts, thus spanning the entire structure (Figs. 5, 6). The four contiguous $\beta$-strands 3, 4, 5, and 6 represent a topology that has not previously been observed (Zhang & Kim, 2000). Comparison between the YrdC fold and those of all other protein structures currently deposited in the Protein Data Bank (PDB) using the program DALI (Holm & Sander, 1993) confirmed that YrdC adopts a new fold. The structural alignments yielded a Z-value of 1.8 (RMSD 4.9 Å) between a 55-amino acid fragment of translation initiation factor 3 (PDB code 1tig) and YrdC, but pairs with a $Z < 2.0$ can be regarded as structurally dissimilar. Although open $\alpha/\beta$-sheet structures are quite common, they also vary considerably in their size and strand order (Branden & Tooze, 1999). It appears that the order of $\beta$-strands in the YrdC structure is unique. A further remarkable feature of this structure is the extreme twisting between some of the adjacent $\beta$-strands (Fig. 5B).

Sequence alignments and locations of conserved residues in the 3D model

Comparison of the amino acid sequences of the YrdC/YciO, Sua5, and HypF families of proteins reveals amino acids that are conserved among all or most of the individual family members, but also distinct differences between the three families (Fig. 7). For example, basic residues with a high degree of conservation in the three families include K50, R52, and R188 (the numbering refers to the sequence of YrdC from E. coli, first line Fig. 7). In addition, S139 and N141 are highly conserved in all three families, as are a number of hydrophobic residues and four prolines. An interesting feature revealed by the alignments is the absence of any conserved acidic amino acids.

There are several sequence patterns that are only conserved in one or two among the three protein families. For example, the sequence P26, T27, and D/E28 is strictly conserved in the YrdC/YciO and Sua5 families, but is not present in the HypF protein family. Instead, the corresponding conserved residues in the latter family are G26, N27, and D/E28.

Fig. 4. Overall fold of the YrdC protein: Stereo diagram of the $\alpha$-carbon backbone with every 10th residue numbered. The drawing was generated in the INSIGHT suite. The view is similar to that in Figures 3 and 5A.

Fig. 5. Schematics of the fold of the YrdC protein drawn with the program RIBBONS (Carson, 1997). A: View approximately along $\beta$-strand 6 in the core of the protein. B: View of the protein rotated by 90° around the vertical relative to the orientation in A, revealing a spoke-like arrangement of $\beta$-strands with the $\alpha$-helical regions situated on the exterior. The $\alpha$-helices and $\beta$-strands are colored cyan and green, respectively, and are labeled with numbers in roman and italic fonts, respectively. Loop regions are colored orange and N- and C-termini are labeled.
family are K, G, and I (L, V) (Fig. 7). This region is part of a loop on one side of the putative binding surface (vide infra) and may indicate functional differences between the protein families. Other regions in the three protein families that exhibit distinct patterns of conservation include amino acids 117, 175, and 176. In the YrdC/YciO and Sua5 families, amino acid 117 is always arginine. In the case of positions 175 and 176, the YrdC and YciO proteins and their homologs exhibit no conservation, while the corresponding conserved pair of amino acids in the Sua5 family of proteins is serine and threonine. In the HypF family, amino acid 174 is also conserved and the conserved trio of amino acids is DDS (Fig. 7). Again, these residues are on the putative binding surface and are likely to reflect differences in the functions of the three families.

A further region with different degrees of conservation and differences among conserved residues between the three families concerns amino acids 189–192. Amino acid 190 is always glycine in the YrdC/YciO and Sua5 protein families, with Pro189 being conserved in addition to the latter family. By comparison, in the HypF family, amino acids 189 and 190 are not conserved; instead residues 191 and 192 are always arginine and glycine, respectively, in that family. These sequence patterns suggest that the structure observed for the E. coli YrdC protein will probably also be adopted by the homologous domains in the Sua5 and HypF families of proteins. However, proteins belonging to the three families may display structural differences at the local level. This region and that following it in the multidomain proteins is likely to represent an interdomain linker, potentially going along with specialized functions.

In the 3D structure of E. coli YrdC, the majority of conserved amino acids discussed above lie on or near the surface of the protein. Most of them map to the floor of a depression constituted by the highly twisted β-sheet and flanked on one side by a ridge formed by the carboxy terminus of helix 2 and the adjacent loop as well as the loop between the single-turn α5 and the β5 strand (Figs. 5B, 8). The C-terminal 20 residues flank the concave surface on the opposite side of the above ridge (Fig. 9A). The orientation of the protein is the same as in Figs. 5 and 8). When the C-terminus is removed, the shape of the YrdC protein takes on the appearance of a baseball glove (Fig. 9B). In particular, all conserved basic residues are located within this concave surface; K50 lies at the carboxy terminus of α2, R52 is part of the adjacent loop (Fig. 7), and both are arranged beneath the “thumb” of the baseball glove (Fig. 8). Other conserved residues are clustered in close vicinity of the two basic residues, including T25, E28, R117, S139, and N141. In the crystal structure, the strictly conserved R188 is situated at some distance from this cluster of conserved residues, above the concave surface. However, as mentioned above, the orientation of the C-terminal portion of YrdC may be affected by particular packing interactions in the crystal and thus may be somewhat arbitrary. It is likely that the arm formed by residues 170–190 can swing away from the concave surface, thus leading to an altered relative orientation between R188 and the other conserved residues located inside the depression.

When the E. coli YrdC and YciO proteins and 13 of their homologs are aligned separately (data not shown), several additional residues, such as D152, E153, and D180, also exhibit a high degree of conservation. Although they are located on the same face of the YrdC structure as the depression (Fig. 8), they map to sites that are located outside the concave surface and the bordering ridge. The particular arrangement of residues clustered inside the concave surface and in a more scattered fashion on the same face of the protein as well as in a potentially flexible arm (C-terminal region) appears to argue against an enzymatic role of YrdC. Rather, the distribution of conserved residues on the YrdC surface is suggestive of a binding function and the dimension of the depression is consistent with a relatively large ligand.

**Electrostatic surface potential**

Computation of the electrostatic surface potential for YrdC using the program GRASP (Nicholls et al., 1993) reveals that opposite faces of the protein exhibit contrasting polarities. The concave surface and some of the regions protruding from the adjacent ridge feature a positive electrostatic potential (Fig. 9A, B), whereas the potential on the opposite face with an overall convex shape is negative (Fig. 9C). It is noteworthy that the positive charge in the palm of the glove is more or less evenly distributed, with a somewhat weaker negative potential extending to the region buried underneath the C-terminal arm comprising residues 170–190 (Fig. 9B).

By contrast, other areas surrounding the depression display relatively strong negative potentials. A search for proteins with shape and electrostatic surface potential comparable to the concave surface of YrdC yielded at least one example (Fig. 9A, D). The structure of the signal recognition particle Alu RNA binding heterodimer SRP 9/14 (Birse et al., 1997) features a more or less circular depression with a diameter similar to that of YrdC (~20 Å). In addition, the curvatures of the positively charged surfaces are fairly similar in the two cases and patches with negative electrostatic potential surrounding the depression constitute a further shared feature between the two proteins. As with YrdC, the structural scaffold underlying the concave surface of the SRP 9/14 heterodimer is made up of a highly twisted β-sheet. However, in the case of the latter, the sheet contains the dimer interface and both constituting proteins thus contribute β-strands.

**Nucleic acid binding studies**

In view of the dimensions and positive electrostatic potential of the YrdC binding surface as well as its resemblance to that of the SRP 9/14 heterodimer, we turned to nucleic acids as potential ligands of YrdC. For a quantitative estimation of the binding efficiencies between YrdC and a variety of nucleic acid fragments, protein fluorescence intensity at 345 nm (excitation at 295 nm) was measured. The relative intensity I/I₀ was then plotted as a function of
Fig. 7. Sequence alignments of the YrdC/YciO, Sua5, and HypF families of proteins from a variety of organisms labeled on the left-hand side. Conserved residues are highlighted: Blue, basic; yellow, hydrophobic; gray, Gly and Pro; magenta, all others. Boxed residues highlight selected regions where sequence conservation differs between the three protein families. Secondary structure elements based on the crystal structure of YrdC from *E. coli* are indicated above the sequences — cyan, α-helices; green, β-strands. The numbering refers to the YrdC protein from *E. coli*. The NCBI genbank ID numbers of the included sequences are: *E. coli* yrdC | gi:2851671; *Deinococcus radioduran* | gi:6459642; *Caenorhabditis elegans* | gi:5832914; *Aquifex aeolicus* | gi:2983259; *E. coli* | gi:3915969; *Streptomyces coelicolor* | gi:5688851; *Synechocystis* | gi:3024006; *Arabidopsis thaliana* | gi:4181799; *Methanococcus jannaschii* | gi:3024906; *Pyrococcus abyssi* | gi:5458658; *Methanobacterium thermoautotrophicum* | gi:3234657; and *E. coli* | gi:2506575.
oligonucleotide concentration ($I_o$ and $I$ are the fluorescence intensities of YrdC alone and the YrdC-oligonucleotide complex, respectively) (Fig. 10A). Addition of either an RNA dodecamer duplex with sequence CGCGAAUUGCG (r-12) or a mixture of tRNAs to the protein solution caused about 25% quenching of the intrinsic Trp fluorescence of YrdC. Conversely, additions of a single-stranded RNA with sequence GGAUUAGG (r-10) or a DNA dodecamer duplex with sequence CGCGAATTGCAG (d-12) had almost no effect on protein fluorescence in the corresponding concentration region (less than 5% change). The fluorometric titration data allow evaluation of the equilibrium dissociation constants for binding of r-12 and tRNA by YrdC. The titration curves were simulated assuming a 1:1 stoichiometry of the complex (single strands in the case of r-10 and tRNA and duplexes in the case of d-12 and r-12). Dissociation constants and $f$ values that resulted in optimal fits are presented in Table 3. The estimated binding affinities for RNA dodecamer duplex and tRNA are rather high ($K_D \sim 10^{-7} \text{ M}$), suggesting a specificity of YrdC for RNAs featuring double stranded portions.

**Functional implications**

The observation that *E. coli* YrdC exhibits selective binding of RNA over DNA is consistent with a role of the protein in translation. This conclusion is also reasonable considering the ability of yeast Sua5, a protein with a domain that is homologous to YrdC,
with a host of other RNA-binding proteins. The structural analysis of YrdC also provides evidence that it is the YrdC-like domain of Sua5 that is involved in RNA binding. Obviously the structure of YrdC alone does not shed any light on the precise manner by which Sua5 suppresses an aberrant message and arguing against its involvement at the level of transcription.

Table 3. Estimated binding affinities of YrdC for duplex RNA and transfer RNA

<table>
<thead>
<tr>
<th>Nucleic acid</th>
<th>$K_D$ (μM)$^b$</th>
<th>$f^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>[${\text{CGCGAAUUCGCG}}$]_2</td>
<td>0.18 ± 0.04</td>
<td>0.75 ± 0.01</td>
</tr>
<tr>
<td>tRNA$^d$</td>
<td>0.68 ± 0.15</td>
<td>0.75 ± 0.02</td>
</tr>
</tbody>
</table>

$^a$Fluorescence intensities $I_0$ and $I$ of protein and protein-nucleic acid complex, respectively, are related by $(P_0 + N_0)$ are the total concentrations of protein and nucleic acid, respectively; Kurganov et al., 1972: $I/I_0 = 1 + (f - 1) + (((K + P_0 + N_0) - \sqrt{[(K + P_0 + N_0)^2 - (4 + P_0 + N_0)]/2 + P_0})$.

Dissociation constant $K_D$ and fluorescence intensities ratio $f$ of bound and free protein are the results of fitting theoretical curves according to the above formula to the experimental data points.

$bK_D = [P]/[N]/[PN]; [P]$ and $[N]$ are the concentrations of protein and RNA, respectively.

$cRatio of fluorescence intensities of bound and free protein.

dMixture of transfer RNAs (see Materials and methods).

Materials and methods

Overexpression and purification of the wt and Se-Met YrdC

The yrdC gene was cloned into the pET-29b vector, which was then transformed into the $E. coli$ BL21(DE3) strain. Overexpression of the His-tagged fusion protein (amino acids 1–192 and a six-residue his-tag) was induced with isopropyl-$\beta$-D-thiogalactoside (IPTG). Cells suspended in 250 mM NaCl, 30 mM Tris-HCl, pH 8.0, and 0.1 mM EDTA were lysed with 50 mg/mL lysozyme and 10 U/mL DNase I over the course of 30 min ($RT$). The protein was purified by affinity chromatography using a Hi Trap column (Pharmacia Biotech, Piscataway, New Jersey) charged with Ni$^{2+}$.

The yield was about 10 mg pure protein per 3 g of cells. Selenomethionine (Se-Met) YrdC was prepared in a similar fashion using a standard protocol to saturate the biosynthetic pathway for methionine production (Doublié, 1997).

Crystallization and data acquisition

Crystallization conditions were screened by the sparse matrix technique (Jancarik & Kim, 1991), using the Hampton Research
Structure determination and refinement

The structure of the YrdC protein was determined by the MAD technique. Two Se sites were located in each of the two molecules that constitute the asymmetric unit and MAD phases were calculated with the program CNS-solve (Brünger, 1998). Solvent flipping (Abrahams & Leslie, 1996) and NCS (noncrystallographic symmetry restraints) density averaging significantly improved the initial phases and electron density maps allowed tracing of the entire protein C-alpha chain with the exception of the first three residues and six residues in the flexible loop region near the C-terminus. A model was built with the program O (Jones & Kjeldgaard, 1997) and refined with the program CNS (Brünger, 1998) using positional, simulated annealing, and temperature factor refinement cycles. The free R-factor was monitored by setting aside 10% of the reflections as a test set (Brünger, 1992). The missing first three residues and the six residues near the C-terminus were added at this point, but the His-tag was not visible in the electron density maps. Only weak electron density was present for residues near the C-terminus in both subunits. The final geometries of the C-terminal regions of the two subunits were modeled separately, and no symmetry constraints were used. Along with two YrdC molecules, 386 water molecules and two phosphate ions were built into the electron density map. The Ramachandran plot calculated with the program PROCHECK (CCP4, 1994) indicates that 100% of the nonglycine and nonproline residues in the final model lie in the most favored and additional allowed regions. Final refinement parameters and RMSDs from standard geometry are listed in Table 2.

Sequence alignments and structure comparisons

An updated list of the clusters of orthologous groups (Tatusov et al., 1997, 2000) can be found at the following website: http://www.ncbi.nlm.nih.gov/COG/. All sequence searches were performed with PSI-BLAST (Altschul et al., 1997) at the National Center for Biotechnology (NCBI) website (http://www.ncbi.nlm.nih.gov/BLAST). CLUSTALX was used for the sequence alignment in Figure 7 (Jeanmougin et al., 1998). Comparisons between the YrdC structure and those of other proteins were performed with DALI (Holm & Sander, 1993) at the DALI website (http://www.ebi.ac.uk/dali/).

Fluorescence measurements

The YrdC protein contains tryptophans at positions 89 and 106, rendering it suitable for fluorescence studies. The fluorescence characteristics of native YrdC (wavelength spectral maximum 325 ± 1 nm, spectral band 40 ± 1 nm, and quantum yield 0.103 ± 0.01) indicate that tryptophan residues exhibit relatively low mobility and are localized in regions that are inaccessible to bulk solvent. The DNA dodemecer was synthesized by the phosphoramidite method (Pharmacia Expeditite). RNA oligonucleotides were purchased from Dharmacon (Boulder, Colorado), and all oligomers were deprotected and purified by reversed phase high-performance liquid chromatography. The transfer RNA mixture was purchased from Sigma (St. Louis, Missouri) (E. coli strain W, type XXI, Sigma R4251). Steady-state techniques were employed to monitor changes in the fluorescence intensity of W89 and W106 upon addition of nucleic acid molecules. Titrations of the protein with DNAs and RNAs were performed while monitoring the change of the intrinsic tryptophan fluorescence of YrdC (λem = 295 nm; λex = 345 nm), using a Shimadzu 5000 (Japan) spectrophotometer. The temperature of the cell compartment was kept at 20 ± 0.1°C, using a refrigerated circulating water bath. All measurements were corrected for dilution and blank solutions. Because the contribution to the absorbance at 295 nm by the oligonucleotides was below 0.04 at the highest concentration used, no inner filter correction was needed. Dissociation constants were calculated with the program ORIGIN (Table 3).

References

Cambillau C, Roussel A. 1997. Turbo frodo, version OpenGL I. Marseille, France: Université Aix-Marseille II.

Acknowledgments

We are grateful to Professor B.I. Kurganov for discussions and help with calculation of the equilibrium constants. This work was partly supported by NIH Grant GM-55237 (to M.E.). Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract W-31-109-Eng-38.


