Protein Science (2000), *9*:2557–2566. Cambridge University Press. Printed in the USA.

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

MARIANNA TEPLOVA,[1] VALENTINA TERESHKO,[1] RUSLAN SANISHVILI,[2] ANDRZEJ JOACHIMIAK,[2] TATYANA BUSHUEVA,[3] WAYNE F. ANDERSON,[4] AND MARTIN EGLI[1]

[1]Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee 37235
[2]Structural Biology Center, Biosciences Division, Argonne National Laboratory, Argonne, Illinois 60439
[3]Cardiology Research Center, Russian Academy of Medical Sciences, Moscow 121552, Russia
[4]Department 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 11*: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 protein 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.

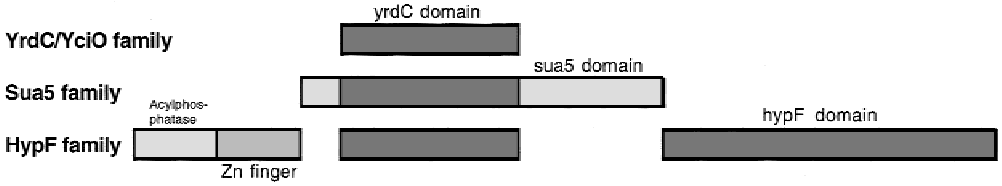

Fig. 1. Schematic of the YrdC and YciO families of proteins and the locations of yrdC domains within the sua5 and hypF families of proteins. YrdC from *E. coli* comprises 192 amino acids, Sua5 from yeast is 420 amino acids long, and the length of the HypF family ranges between 700 and 800 amino acids.

homologous to *sua5* are yet available expressed sequence tags (ESTs) indicate that a human homolog exists.

A regulatory role has also been ascribed to the other family of proteins featuring a YrdC domain, HypF (Fig. 1). In *E. coli*, the HypF protein is involved in H_2 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 H_2 (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 HycI 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 HypF protein family (Fig. 1). A further characteristic of this family is a putative acylphosphatase domain at the N-terminus (Wolf 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; Volz, 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)*

Wavelength (Å)	Number of reflections		Completeness (%)	R_{sym} (%)[a]	Phasing power λ_3/λ_i
	Measured	Unique	All (last shell)	All (last shell)	(Friedel)
λ_1 = 0.9795 (edge)	231,530	62,907	96.2 (99.5)	6.0 (42.3)	2.45 (3.34)
λ_2 = 0.9794 (peak)	230,966	62,780	96.2 (99.8)	6.2 (48.6)	1.55 (3.05)
λ_3 = 0.9465 (reference)	229,370	61,692	94.7 (100.0)	5.3 (35.4)	— (1.76)

Overall figure of merit: 0.77

[a] $R_{sym} = \sum_{hkl}\sum_i|I(hkl)_i - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i\langle I(hkl)_i\rangle$ for the intensity I of i observations of a reflection hkl.

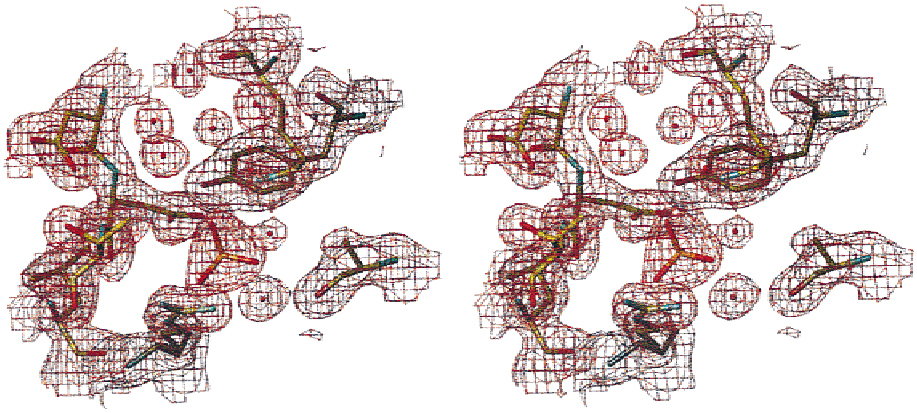

Fig. 2. Quality of the final model of the *E. coli* YrdC crystal structure. Stereo diagram of the $(2F_o - F_c)$ Fourier sum electron density around the phosphate ion bound to K69, D74, and R82 in one of the YrdC subunits, drawn with TURBO FRODO (Cambillau & Roussel, 1997). The map is contoured at the 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.

C-terminal amino acids (8 in the case of subunit 2), the His tag is invisible. An example of the $(2F_o - F_c)$ 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 α-carbon atoms in their backbones is 0.60 Å. The greatest conformational variation is found at their C-termini and involves residues G170–F187 (RMSD 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 × 30 × 30 Å. A stereo diagram of the α-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 α/β twisted open-sheet structure with parallel and antiparallel orientations of adjacent β-strands. The structure comprises seven α-helices and seven β-strands with an unusual order and topology of β-strands and very strong twisting between them. An α-helical region at the N-terminus of the protein (α1; Fig. 5) is followed by two antiparallel β-strands (β1 and β2). These lead over to the second helix α2 that is then linked to β3 by a five-residue loop. The third β-strand is located near the center of the protein, and a short turn leads over to α3 and α4 that form the other end of the protein together with β4. An irregular region interrupted only by the single-turn helix α5 provides a connection to β5, itself oriented parallel to β3 and antiparallel to β4. The loops at the carboxy ends of β3 and β5 connect them to α-helices on opposite sides of the sheet (α3 and α6, respectively; Fig. 5). Thus, α6 is inserted between α3 and α1 and leads back to the center of the molecule where β6 is inserted antiparallel to β3. A 12-residue loop links β6 to α7 and a long loop interrupted by the short β7 leads to the C-terminus.

Considerable conformational flexibility of the C-terminal region is indicated by high temperature factors of residues there (46.0 Å^2 on average for G170 to R188, subunit 1; 42.6 Å^2 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 2. *Refinement statistics*

Parameter	Value
Resolution (Å)	25.0–2.0 (bulk solvent correction)
Reflections ($F \geq 2\sigma[F]$)	51,216
R-factor (%)	20.2
R-free (%)	21.3
Number of atoms	3,244 (2 YrdC, 2 phosphates, 386 H_2O)
RMSDs	
Bond length (Å)	0.009
Bond angles (°)	1.5
Average temperature factors	
Protein atoms (Å^2)	25.4 (subunit 1), 23.8 (subunit 2)
Solvent molecules (Å^2)	34.2
Phosphate ion	49.5

[a]RMSDs from ideal geometry.

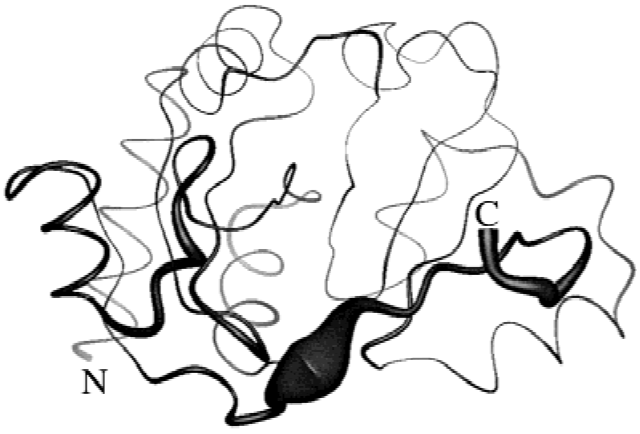

Fig. 3. Conformational variations between the two independent YrdC molecules per crystallographic asymmetric unit, drawn with the program MOLMOL (Koradi et al., 1996). The thickness of the line is proportional to the distance between corresponding α-carbon atoms of individual molecules. The largest deviations are found between residues R172 and P175.

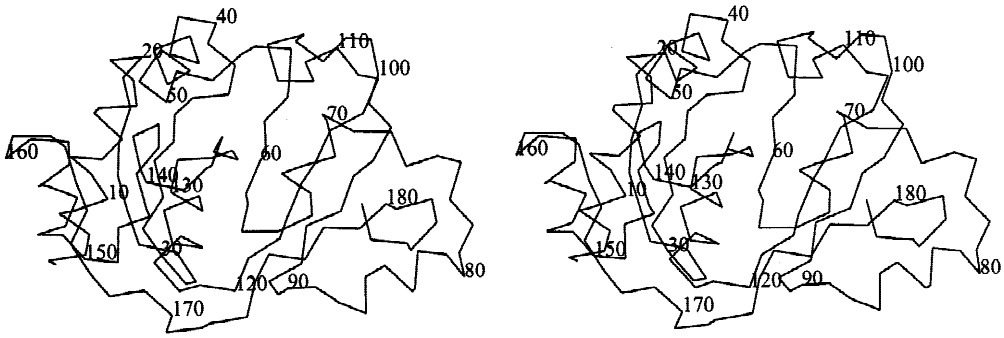

Fig. 4. Overall fold of the YrdC protein: Stereo diagram of the α-carbon backbone with every 10th residue numbered. The drawing was generated in the INSIGHT suite. The view is similar to that in Figures 3 and 5A.

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 (F187 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 α1, β1, β2, and α2, and the second comprises β3, α3, α4, β4, α5, and β5. These two portions are bridged by α6 and β6 that constitutes the third part. The adjacent flexible C-terminal region including α7 and β7 is folded down over all three parts, thus spanning the entire structure (Figs. 5, 6). The four contiguous β-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 α/β-sheet structures are quite common, they also vary considerably in their size and strand order (Branden & Tooze, 1999). It appears that the order of β-strands in the YrdC structure is unique. A further remarkable feature of this structure is the extreme twisting between some of the adjacent β-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

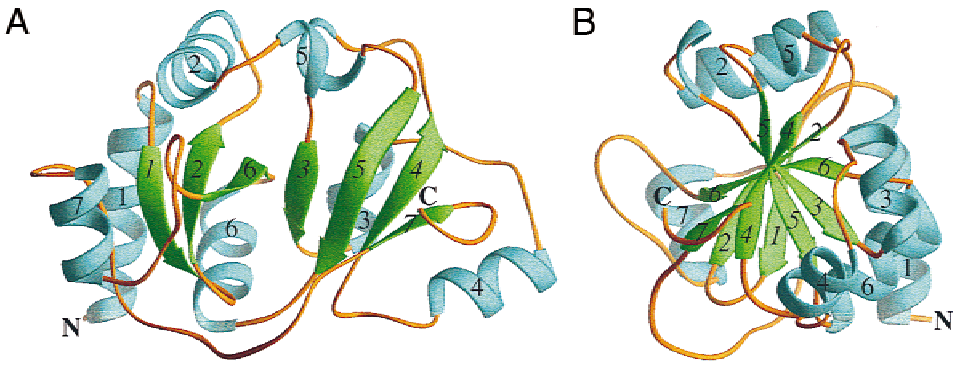

Fig. 5. Schematics of the fold of the YrdC protein drawn with the program RIBBONS (Carson, 1997). **A:** View approximately along β-strand 6 in the core of the protein. **B:** View of the protein rotated by 90° around the vertical relative to the orientation in **A**, revealing a spoke-like arrangement of β-strands with the α-helical regions situated on the exterior. The α-helices and β-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.

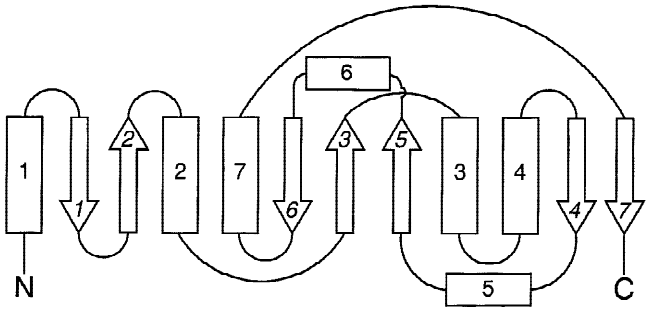

Fig. 6. Cartoon of the folding topology of YrdC. Helices are drawn as boxes, β-strands are drawn as open arrows and are numbered and N- and C-termini are labeled.

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 in 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 positive 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/I_o was then plotted as a function of

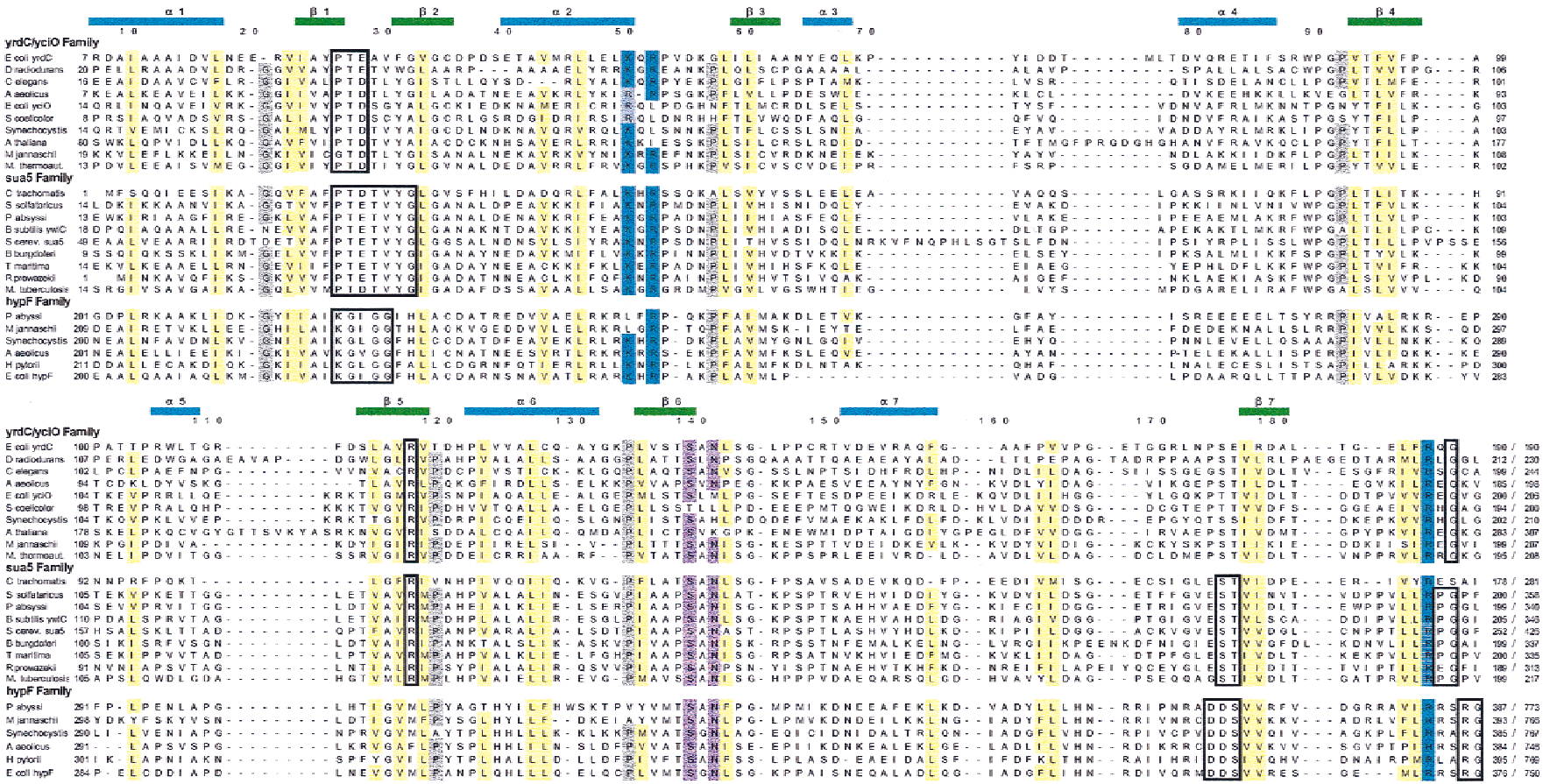

Fig. 7. Sequence alignments of the YrdC/YciO, Sua5, and HypF families of proteins from a variety of organisms labeled on the left-hand side. Conserved residues are highlighted: Blue, basic; yellow, hydrophobic; gray, Gly and Pro; magenta, all others. Boxed residues highlight selected regions where sequence 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), and the numbering refers to the YrdC protein from *E. coli*. The NCBI genbank ID numbers (gi) 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|1651805|; *Arabidopsis thaliana* gi|6091731|; *Methanococcus jannaschii* gi|2500922|; *Methanobacterium thermoautotrophicum* gi|2622823|; *Chlamydia trachomatis* gi|3328537|; *Sulfolobus solfataricus* gi|6015811|; *Pyrococcus absyssi* gi|5459014|; *Bacillus subtilis* gi|732384|; *Saccharomyces cerevisiae* gi|6321269|; *Borrelia burgdorferi* gi|2688669|; *Thermotoga maritima* gi|4981386|; *Rickettsia prowazekii* gi|3861373|; *Mycobacterium tuberculosis* gi|1722959|; *Pyrococcus abyssi* gi|5458658|; *Methanococcus jannaschii* gi|3024906|; *Synechocystis* gi|3024006|; *Aquifex aeolicus* gi|2983269|; *Helicobacter pylori* gi|2313122|; and *E. coli* gi|2506575|.

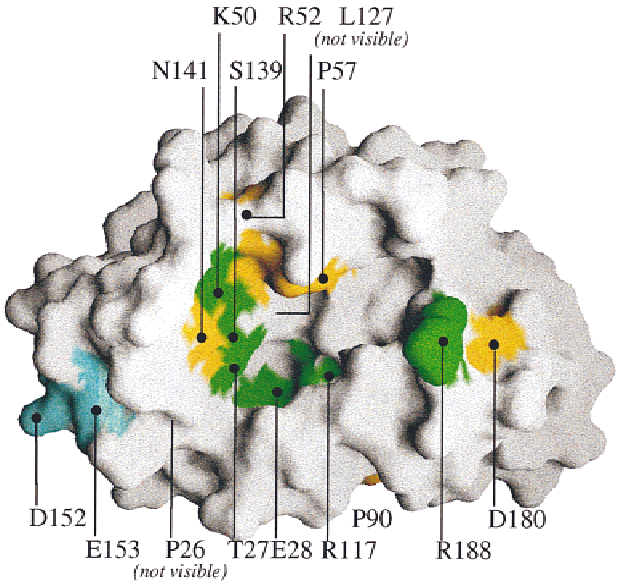

Fig. 8. Arrangement of conserved residues on the 3D surface of YrdC. Green, yellow, and cyan patches indicate the locations of amino acids that are conserved in 15 or more, in 13 or 14 out of 16, and in 12 out of 16 homologs of YrdC analyzed, respectively (10 of which are depicted in Fig. 7), and all other residues are white. Residues A42, G132, and P134 all constitute yellow patches on the backside of the protein. The view is identical to the one in Figures 3 and 4A.

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 CGCGAAUUCGCG (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 GGAUUAUAGG (r-10) or a DNA dodecamer duplex with sequence CGCGAATTCGCG (d-12) had almost no effect on protein fluorescence in the corresponding concentration region (less then 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 ($K_D \sim 10^{-7}$ 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,

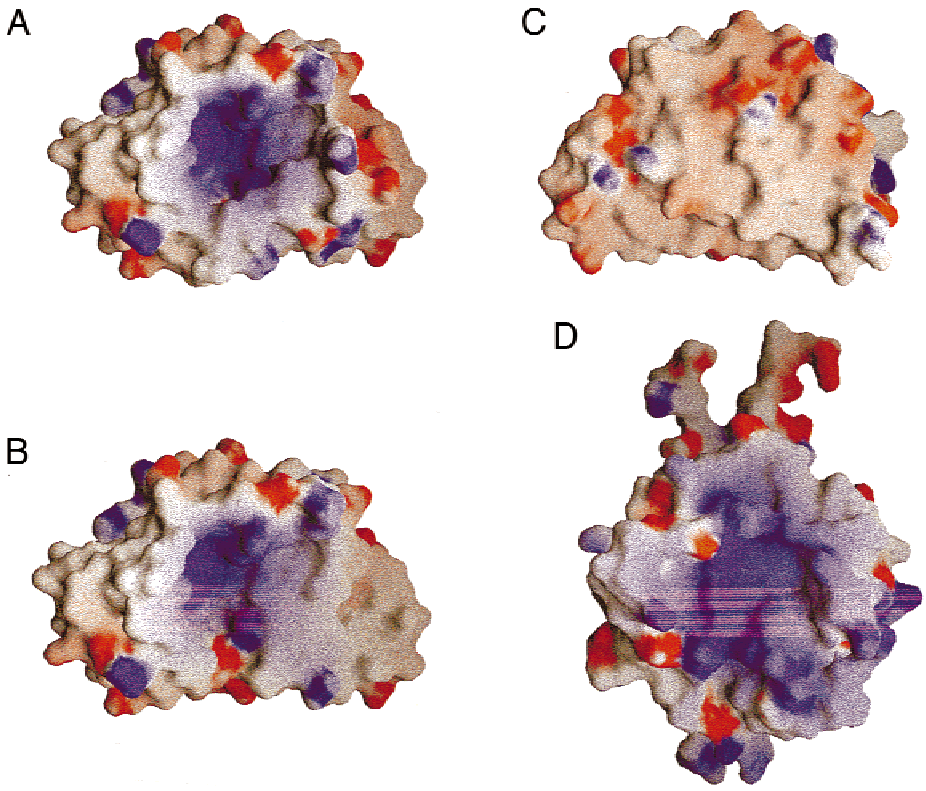

Fig. 9. Electrostatic surface potential of YrdC calculated with the program GRASP (Nicholls et al., 1993). **A:** View of the protein identical to the one in Figures 3, 4A, and 8, with the C-terminal 20 residues including $\beta 7$ visible in the foreground. **B:** The same orientation as in **A** but with the C-terminal 20 residues removed. **C:** View of the protein rotated around the vertical by 180° relative to the orientation in **A**. **D:** Electrostatic surface potential representation of the signal recognition particle *Alu* RNA binding heterodimer SRP9/14 from mouse, based on a crystal structure determination (Birse et al., 1997).

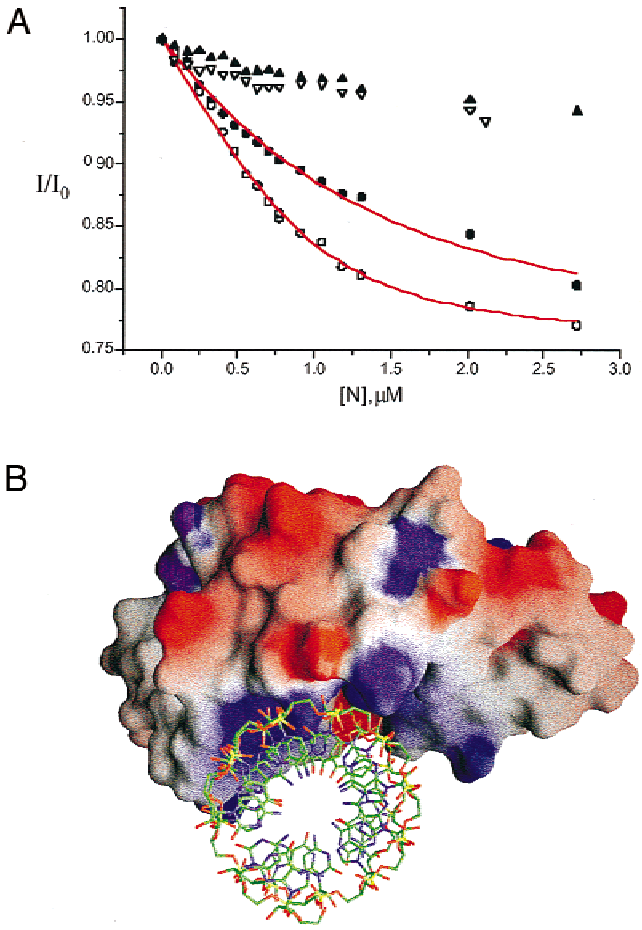

Fig. 10. A: Relative Trp fluorescence intensity for YrdC as a function of oligonucleotide concentration (▲, d-12; ▽, r-10; ●, tRNA; ○, r-12). The concentration of YrdC in all experiments was 1 μM, using the following conditions: 25 mM Tris-HCl, pH 8.0, 100 mM NaCl, 20°C. Solid lines indicate theoretical values of the fluorescence intensity ratio computed according to the difference method (Kurganov et al., 1972) (Table 3). **B:** Model of a complex between *E. coli* YrdC and an RNA dodecamer duplex, illustrating the complementary shapes and the close spacing between several basic protein side chains and RNA phosphate groups. To generate the complex, the flexible region comprising the C-terminal 20 residues of YrdC was removed. The protein is drawn as a surface model, and the blue patches indicate areas of positive electrostatic potential. The RNA is drawn as a stick model and atoms are colored green, red, blue, and yellow for carbon, oxygen, nitrogen, and phosphorus, respectively.

to suppress a translation initiation defect in the leader region of the *cyc1* gene. Although we can demonstrate a preference of YrdC for RNAs comprising double-stranded regions relative to single-stranded species, the precise nature of the RNA target(s) of YrdC remains to be determined. A standard A-form RNA helix can be docked into the concave surface of YrdC in a straightforward manner (Fig. 10B). However, the RNAs actually recognized by YrdC or Sua5 will likely exhibit additional secondary and tertiary structural components besides simple canonical stem regions. Taking into account that the dissociation constants of complexes between YrdC and an RNA duplex and tRNAs are in the 200–700 nM range, the dissociation constant for a complex between YrdC and its natural target might be expected to lie in the low nanomolar range, in accordance with the specificities observed

Table 3. *Estimated binding affinities of YrdC for duplex RNA and transfer RNA*[a]

Nucleic acid	K_D (μM)[b]	f[c]
$[r(CGCGAAUUCGCG)]_2$	0.18 ± 0.04	0.75 ± 0.01
tRNA[d]	0.68 ± 0.15	0.75 ± 0.02

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

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.

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

[c] Ratio of fluorescence intensities of bound and free protein.

[d] Mixture of transfer RNAs (see Materials and methods).

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 *cyc1* initiation codon. Thus, the protein could mediate interactions between the *cyc1* message and the ribosome or, alternatively, could stabilize the messenger RNA. In this regard, it is of interest to determine whether Sua5 binds its own message. To summarize, we conclude that the YrdC protein from *E. coli* that occurs as a domain in the Sua5 and HypF families of proteins exhibits a new fold. The YrdC protein features a large concave surface with an overall positive electrostatic potential. The protein binds double-stranded RNA with high specificity, providing support for a role of yeast Sua5 as a translational facilitator and arguing against its involvement at the level of transcription.

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-β-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

(Laguna Niguel, California) Crystal Screen I. Equal volumes of aqueous protein solution (10 mg/mL) in 24 mM Tris-HCl, pH 7.5, 5 mM DTT, and crystallization buffer were mixed and sitting drops were equilibrated against 0.5 mL reservoirs. Very thin crystals with size 0.05×0.2 mm grew within one week from a solution containing 5 mg/mL YrdC, 0.6 M sodium potassium phosphate, pH 5.5, and after seeding repeatedly. The space group of the crystals is monoclinic $P2_1$ with cell constants $a = 60.50$ Å, $b = 71.66$ Å, $c = 55.26$ Å, and $\beta = 103.2°$, and two molecules per asymmetric unit. For data collection, a Se-Met crystal was flash frozen (100 K) in mother liquor that was supplemented with 28% sucrose. Diffraction data to a maximum resolution of 1.96 Å were collected at three wavelengths on the 19-ID beamline of the Structural Biology Center at the Advanced Photon Source (Argonne, Illinois), using a 3×3 mosaic CCD detector (Table 1). Data were integrated and merged in the HKL 2000 suite (Otwinowski & Minor, 1997).

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 dodecamer was synthesized by the phosphoramidite method (Pharmacia Expedite), 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 ($\lambda_{ex} = 295$ nm; $\lambda_{em} = 345$ nm), using a Shimadzu 5000 (Japan) spectrofluorometer. 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).

Coordinates

Coordinates and structure factors have been deposited in the PDB (ID code 1HRU).

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.

References

Abrahams JP, Leslie AGW. 1996. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr D 52*:30–42.

Altschul SF, Madden TL, Schäffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res 25*:3389–3402.

Birse DEA, Kapp U, Strub K, Cusack S, Åberg A. 1997. The crystal structure of the signal recognition particle *Alu* RNA binding heterodimer, SRP9/14. *EMBO J 16*:3757–3766.

Branden C, Tooze J. 1999. *Introduction to protein structure*, 2nd ed. New York: Garland Publishing Inc.

Brünger AT. 1992. Free *R* value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature 355*:472–475.

Brünger AT. 1998. *Crystallography & NMR system (CNS)*, version 0.4. New Haven, Connecticut: Yale University.

Cambillau C, Roussel A. 1997. *Turbo frodo*, version OpenGL.1. Marseille, France: Université Aix-Marseille II.

Carson M. 1997. RIBBONS. *Methods Enzymol 277*:493–505.

CCP4 (Collaborative Computational Project Number 4). 1994. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D 50*:760–763.

Colbeau A, Elsen S, Tomiyama M, Zorin NA, Dimon B, Vignais PM. 1998. *Rhodobacter capsulatus* HypF is involved in regulation of hydrogenase synthesis through the HupUV proteins. *Eur J Biochem 251*:65–71.

Colbeau A, Richaud P, Toussaint B, Caballero FJ, Elster C, Delphin C, Smith RL, Chabert J, Vignais PM. 1993. Organization of the genes necessary for

hydrogenase expression in *Rhodobacter capsulatus*. Sequence analysis and identification of two hyp regulatory mutants. *Mol Microbiol 8*:15–29.
Colovos C, Cascio D, Yeates TO. 1998. The 1.8 Å crystal structure of the ycaC gene product from *Escherichia coli* reveals an octameric hydrolase of unknown specificity. *Structure 6*:1329–1337.
Cort JR, Koonin EV, Bash PA, Kennedy MA. 1999. A phylogenetic approach to target selection for structural genomics: Solution structure of YciH. *Nucleic Acids Res 27*:4018–4027.
Doublié S. 1997. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol 276*:523–530.
Drapal N, Bock A. 1998. Interaction of the hydrogenase accessory protein HypC with HycE, the large subunit of *Escherichia coli* hydrogenase 3 during enzyme maturation. *Biochemistry 37*:2941–2948.
Hampsey M, Na JG, Pinto I, Ware DE, Berroteran RW. 1991. Extragenic suppressors of a translation initiation defect in the *cyc1* gene of *Saccharomyces cerevisiae*. *Biochimie 73*:1445–1455.
Holm L, Sander C. 1993. Protein structure comparison by alignment of distance matrices. *J Mol Biol 233*:123–128.
Hwang KY, Chung JH, Kim S-H, Han YS, Cho Y. 1999. Structure-based identification of a novel NTPase from *Methanococcus jannaschii*. *Nat Struct Biol 6*:691–696.
Jancarik J, Kim S-H. 1991. Sparse matrix sampling: A screening method for crystallization of proteins. *J Appl Crystallogr 24*:409–411.
Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ. 1998. Multiple sequence alignment with ClustalX. *Trends Biochem Sci 23*:403–405.
Jones TA, Kjeldgaard M. 1997. Electron-density map interpretation. *Methods Enzymol 277*:173–208.
Kim SH. 1998. Shining a light on structural genomics. *Nat Struct Biol 5*:643–645.
Koradi R, Billeter M, Wüthrich K. 1996. MOLMOL, a program for display and analysis of macromolecular structures. *J Mol Graphics 14*:51–55.
Kurganov BI, Sugrobova NP, Yakovlev VA. 1972. Estimation of dissociation constant of enzyme–ligand complex from fluorometric data by "difference" method. *FEBS Lett 19*:308–310.
Maier T, Binder U, Bock A. 1996. Analysis of the hydA locus of *Escherichia coli*: Two genes (hydN and hypF) involved in formate and hydrogen metabolism. *Arch Microbiol 165*:333–341.
Minasov G, Teplova M, Stewart GC, Koonin EV, Anderson WF, Egli M. 2000. Functional implications from the structures of the *Bacillus subtilis* protein Maf with and without dUTP. *Proc Natl Acad Sci USA 97*:6328–6333.
Na JG, Pinto I, Hampsey M. 1992. Isolation and characterization of *sua5*, a novel gene required for normal growth in *Saccharomyces cerevisiae*. *Genetics 131*:791–801.
Nicholls A, Bharadwaj R, Honig B. 1993. GRASP: Graphical representation and analysis of surface properties. *Biophys J 64*:166–170.
Olson JW, Maier RJ. 1997. The sequences of hypF, hypC and hypD complete the hyp gene cluster required for hydrogenase activity in *Bradyrhizobium japonicum*. *Gene 199*:93–99.
Orengo CA, Todd AE, Thornton JM. 1999. From protein structure to function. *Curr Opin Struct Biol 9*:374–382.
Otwinowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol 276*:307–326.
Pinto I, Na JG, Sherman F, Hampsey M. 1992. *Cis*-acting and *trans*-acting suppressors of a translation initiation defect at the cyc1 locus of *Saccharomyces cerevisiae*. *Genetics 132*:97–112.
Sinha S, Rappu P, Lange SC, Mäntsälä P, Zalkin H, Smith JL. 1999. Crystal structure of *Bacillus subtilis* YabJ, a purine regulatory protein and member of the highly conserved YjgF familiy. *Proc Natl Acad Sci USA 96*:13074–13079.
Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res 28*:33–36.
Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. *Science 278*:631–637.
Teichmann SA, Chothia C, Gerstein M. 1999. Advances in structural genomics. *Curr Opin Struct Biol 9*:390–399.
Volz K. 1999. A test case for structure-based functional assignment: The 1.2 Å crystal structure of the yjgF gene product from *Escherichia coli*. *Protein Sci 8*:2428–2437.
Wolf I, Buhrke T, Dernedde J, Pohlmann A, Friedrich B. 1998. Duplication of hyp genes involved in maturation of [NiFe] hydrogenase in *Alicaligenes eutrophus* H16. *Arch Microbiol 170*:451–459.
Yang F, Gustafson KR, Boyd MR, Wlodawer A. 1998. Crystal structure of *Escherichia coli* HdeA. *Nat Struct Biol 5*:763–764.
Zarembinski TI, Hung L-W, Mueller-Dieckmann H-J, Kim K-K, Yokota H, Kim R, Kim S-H. 1998. Structure-based assignment of the biochemical function of a hypothetical protein: A test case of structural genomics. *Proc Natl Acad Sci USA 95*:15189–15193.
Zhang C, Kim S-H. 2000. The anatomy of protein β-sheet topology. *J Mol Biol 299*:1075–1089.