Bacterial RibG is an attractive candidate for development of anti-microbial drugs because of its involvement in the riboflavin biosynthesis. The crystal structure of Bacillus subtilis RibG at 2.41 Å resolution displayed a tetrameric ring-like structure with an extensive interface of ~2400 Å²/monomer. The N-terminal deaminase domain belongs to the cytidine deaminase superfamily. A structure-based sequence alignment of a variety of nucleotide deaminases reveals not only the unique signatures in each family member for gene annotation but also putative substrate-interacting residues for RNA-editing deaminases. The strong structural conservation between the C-terminal reductase domain and the pharmacetically important dihydrofolate reductase suggests that the two reductases involved in the riboflavin and folate biosyntheses evolved from a single ancestral gene. Together with the binding of the essential cofactors, zinc ion and NADPH, the structural comparison assists substrate modeling into the active-site cavities allowing identification of specific substrate recognition. Finally, the present structure reveals that the deaminase and the reductase are separate functional domains and that domain fusion is crucial for the enzyme activities through formation of a stable tetrameric structure.

Flavin coenzymes are ubiquitous in all organisms because of their involvements in central metabolic pathways. Plants and many microorganisms obtain the precursor riboflavin by biosynthesis, whereas animals depend on nutritional sources. Numerous pathogenic microorganisms are unable to take up flavins from the environment and hence have evolved the CDA fold into these various deaminases. Typically useful analogues. The available member structures reveal a virtually identical zinc-assisted deamination mechanism with the consensus histidine and cysteines acting as the zinc ligands, whereas the glutamate serves as a proton shuttle (10–14). However, the question of how nature has evolved the CDA fold into these various deaminases requires additional study.

Furthermore, the structural fold of the R domain was predicted by 3D-PSSM (15) to be similar to dihydrofolate reductase (DHFR). DHFR catalyzes the NADPH-utilizing reduction of dihydrofolate to tetrahydrofolate. Many DHFR inhibitors, such as methotrexate, pyrimethamine, and trimethoprim, have long been used clinically in the treatment of cancer, rheumatoid arthritis, malaria, and bacterial and fungal infection (16). Therefore, the R domain may become an important target for new drug design. To gain structural insights into the inhibitor design, substrate specificity, and evolution, we have solved the BsRibG structure at 2.41 Å resolution.

**EXPERIMENTAL PROCEDURES**

**Protein Preparation**—The His-tagged BsRibG was expressed using the pQE30 vector (Qiagen) in E. coli BL21 pLysS. The recombinant protein contains 12 additional vector residues (MRGSH6GS) at the N terminus. One-liter Luria broth (LB) cultures were grown at 37 °C for 20 h.
galactopyranoside. Cells were grown for another 5 h at 37 °C before harvest. Cell pellets were resuspended in 40 ml of cold buffer A (20 mM Tris-HCl, 200 mM NaCl, pH 7.5) and lysed by French press. After removal of cellular debris by centrifugation at 39,000 g at 4 °C for 30 min, the supernatant was applied to a 5-ml nickel-nitrilotriacetic acid column (Ni-NTA, Qiagen). The resin was washed with 40 mM imidazole in buffer A, and the protein was eluted with buffer A containing 500 mM imidazole. The protein fractions were collected, dialyzed against buffer B (20 mM Tris-HCl, pH 8.5), and loaded onto a Sepharose™ Q column (Amersham Biosciences). After washing with 230 mM NaCl in buffer B, the BsRibG was eluted with 260 mM NaCl and dialyzed against 20 mM HEPES (pH 7.5) and 5 mM dithiothreitol.

### Protein Characterization

The enzyme activity assay was carried out as described previously (4). The substrate was prepared by GTP hydrolysis using the recombinant *E. coli* GTP cyclohydrolase II. The molecular mass in solution was estimated by a Beckman-Coulter XL-A analytical Ultracentrifuge with an An60Ti rotor. Sedimentation velocity was performed at 20 °C and 40,000 rpm with standard double sector centerpieces. The UV absorption of the cells was scanned every 5 min for 2 h, and the data were analyzed using the SedFit program (17).

### Protein Crystallization

The initial crystallization screening was performed with screening kits using the hanging-drop vapor diffusion method at 22 °C. The hanging drops were mixtures of 2 μl of reservoir solution and 2 μl of protein solution. Protein crystals could be obtained under several reservoir solutions, and the best crystals were grown in 26.6% polyethylene glycol 400, 190 mM MgCl₂, 5% glycerol, and 95 mM HEPES (pH 7.5) with a protein solution of 20–25 mg/ml. Crystals appeared and reached their final dimensions in 1 week at 15 °C. The NADPH derivative was prepared by soaking crystals for 4 days in reservoir solution containing 10 mM NADPH. X-ray diffraction data were collected and processed at beamlines BL12B2 at SPring-8 (Harima, Japan) and NW12 at the Photon Factory (Tsukuba, Japan). The crystals belong to the *P*₂₁₂₁ space group with 1 tetramer/asymmetric unit.

### Structure Determination

The BsRibG structure was determined using single-wavelength anomalous dispersion (SAD) of the endogenous zinc ion. Four Zn²⁺ positions were identified and refined with SOLVE (18). The initial phase was improved by direct-method phasing refinement using OASIS (19) before density modification. The initial electron density map showed an interpretable density of ~70% of the tetramer. About 68% of the tetrameric model was automatically built using RESOLVE (18). Because the structural fold was predicted to be similar to yeast cytosine deaminase (yCD) and *Thermotoga maritima* DHFR (TmDHFR) (10, 20), both structures were used to assist in the manual building of the atomic model using TURBO-FRODO (21), and refinement was carried out against the native data to 2.41 Å resolution using crystallography NMR software (22). The statistics for data collection and refinement are summarized in Table 1. About 88% of the residues were in the A, B, C, and D positions of the tetramer.

### Table 1: Statistics for data collection and structural refinement

| Dataset | Native | Zinc-SAD | NADPH |
|---------|--------|----------|--------|
| **Data collection** | | | |
| Wavelength (Å) | 1.0000 | 1.2823 | 1.0000 |
| Unit cell (Å) | 87.3, 108.4, 186.9 | 87.3, 108.5, 187.0 | 86.1, 108.0, 186.8 |
| Resolution range (Å) | 50–2.41 (2.5–2.41) | 50–2.48 (2.57–2.48) | 50–2.97 (3.08–2.97) |
| Total observations | 673,346 (48,412) | 738,284 (55,670) | 313,672 (22,528) |
| Unique reflections | 67,740 (6,370) | 63,545 (5,986) | 35,885 (3,520) |
| Completeness (%) | 97.6 (93.2) | 99.2 (94.9) | 98.2 (98.1) |
| Rmerge (%) | 17.7 (3.9) | 22.9 (5.5) | 18.0 (4.0) |
| **Refinement** | | | |
| Resolution range (Å) | 50–2.41 (2.5–2.41) | 50–2.48 (2.57–2.48) | 50–2.97 (3.08–2.97) |
| Reflections (F > 0) | 67,740 (5,683) | 62,740 (5,683) | 62,740 (5,683) |
| Rcryst (%) for 90% data | 22.6 (28.5) | 22.6 (28.5) | 22.4 (32.4) |
| Rfree (%) for 10% data | 27.1 (31.6) | 27.1 (31.6) | 27.0 (37.5) |
| r.m.s. deviations | 0.009 | 0.007 | 0.007 |
| Bond lengths (Å) | 1.044 | 1.044 | 1.044 |
| Average B-factors (Å²) | 1.44 | 1.31 | 1.31 |
| 10,970 protein atoms | 47.0 | 10,970 protein atoms | 73.0 |
| 4 zinc ions | 42.2 | 4 zinc ions | 77.0 |
| 293 water molecules | 51.8 | 235 water molecules | 52.6 |
| 48 NADPH atoms | 84.5 | | |
idues are in the most favored regions of the Ramachandran plot, with the remaining in the additional allowed regions.

**Sequence Alignment**—To identify the unique signatures for each family member for gene annotation, sequence similarity searches were conducted by PSI-BLAST (23), and multiple sequence alignment of the homologous sequences was performed by ClustalW (24). Because of a conservative hydrophobic core and the consensus HXE and PCX₂⁴⁻⁹C, a structural-based sequence alignment between the CDA members was feasible and was carried out by manual editing according to the available structures. The conserved residues in each family member were mapped onto the known structures to reveal their potential involvement in structural integrity or the enzyme catalysis. Figs. 3, 4, A and B, and 6 were generated by MolScript (25) and Raster3D (26), Fig. 5 by BobScript (25), and Fig. 5B by LigPlot (27).

**RESULTS AND DISCUSSION**

**Overall Structure**—Analytical ultracentrifugation experiments clearly demonstrated that BsRibG exists as a tetramer in solution as well as in crystal form, where the enzyme forms a tetrameric ring-like structure (Figs. 2 and 3a). The current tetrameric model contains residues 1–359 for each subunit. Each monomer is composed of two separate functional domains (Fig. 3b). The N-terminal D domain (residues 1–143) consists of a central five-stranded β-sheet (β₁−β₅) with β₁ running antiparallel to the others. The β-sheet is sandwiched by two helices (α₁A and α₁E) on one side and by three helices (α₂B, α₂C, and α₂D) on the other side. The C-terminal helix α₂F extends away from the D domain and connects to the R domain. The R domain (residues 146–359) is composed of a large nine-stranded β-sheet (β₆A–β₇H and β₈C) with the C-terminal strand β₉H running antiparallel to the others. The β-sheet is flanked by five α-helices (α₁B, α₁C, α₁D, α₁E, and α₁F). The secondary structure elements of the R domain are numbered as for DHFRs (28).

The four subunits in the crystal asymmetric unit did not show significant differences between each individual domain except for several loops (root-mean-square deviations (r.m.s.d.) of 0.38–0.46 and 0.59–0.73 Å for the backbones of residues 2–139 and residues 146–359, respectively). However, the relative orientations between the D and R domains are slightly different, resulting in a weak noncrystallographic symmetry. Molecules A and B interact with each other through their D domains with a buried surface area of ~650 Å² per D domain (the D interface) (Fig. 4A), whereas molecule A makes extensive contacts with molecule C through their respective R domains with a buried area of ~1750 Å² per R domain (the R interface) (Fig. 4B). There are no contacts between molecules A and D in the tetramer.

The D interface is made up mainly of the N-terminal two helices α₁A and α₁B, the β₂ strand, and the connected loops (residues 4–19 and 35–61). There are 14 direct hydrogen bonds between the protein atoms
across the interface and hydrophobic patches formed by Leu8, Leu12, Ile36, Met39, Leu43, and Met57. The anti-parallel disposition of the two pseudodyad-related αA helices also contributes to the interface by dipole-dipole interactions. Interestingly, the side chains of His56 from the two D domains stack very well, with a distance of 3.3–3.4 Å between the aromatic rings. The R interface mainly is made up of: the two large loops, LαA–LαB (residues 156–169) and LαF–LαG (residues 314–338); and the C-terminal residues 339–358. There are 18 direct hydrogen bonds between the protein atoms across the interface, and extensive hydrophobic patches are formed by Try169, Pro314, Lys315, Leu316, Ile317, Leu325, Phe331, Met334, Val337, Leu339, Leu340, Phe342, Ile345, Ile352, and Leu354.

Structural Conservation in the CDA Superfamily—As expected, the D domain displays high structural homology to the available structures of the CDA members including yCD, B. subtilis guanine deaminase (BsGD), CDAs, T4 bacteriophage dCMP deaminase (T4dCMPD), and Aquifex aeolicus TADA (AaTADA), and subdomains 2 and 4 of the chicken AICAR transformylase domain (10–14, 29). The TADAs in prokaryotes and the TAD2/TAD3 heterodimers in eukaryotes are responsible for the I34 alteration at the wobble position of the tRNA anticodon, whereas TAD1 creates the unique 1-methyl-I37 in eukaryotic tRNA (8). Detailed structural comparisons reveal a common three-layer α/β/α structure, in which five β-strands (β1–β5) and three helices (αA–αC) correspond closely, whereas the remainder varies across the different deaminases (Fig. 5A). The main chain atoms of the 65–70 structurally equivalent residues are overlaid with an r.m.s.d. of 0.75 to 1.35 Å and with 8–24% sequence identity. These structural elements in the CDA fold form a conservative hydrophobic core, which is also preserved in the AICAR transformylase domain, implying that the hydrophobic core has been highly conserved throughout evolution.

The active site of the D domain contains one tightly bound endogenous zinc ion, for which the anomalous data provided sufficient phase information for structure determination. The zinc ion is tetrahedrally coordinated by His49 Nε1 (2.0 Å), Cys74 Sγ (2.4 Å), Cys83 Sγ (2.3 Å), and a water molecule (2.0 Å). The zinc-bound water molecule interacts with Glu51 Oγ2 (2.5 Å). The active-site architecture resembles those of the CDA members, which share a similar zinc-assisted deamination mechanism with a virtually identical interaction network between the common moiety of the pyrimidine ring of the substrate, the zinc ion, the zinc-bound water molecule, the zinc ligands, and the base glutamate (Fig. 5B). In addition, the active-site cavities of these deaminases are mainly made up of the C-terminal tail and the loops connecting the αA–αB, α2–αB, β3–αC, and β4–αD (Fig. 5C).

Based on the structural comparison of the active-site cavities, the substrate of the D domain was modeled into the active site through superposition of the nucleobase rings because of the highly conserved interaction networks surrounding the target amino group (Fig. 5B). The model was then subjected to energy minimization with crystallography NMR software. Simulation of the complex structure suggested that the nucleophilic OH-2 group of the pyrimidine ring coordinates to the cat-
competitive zinc ion and interacts with Glu$^{31}$O$^2$ and Cys$^{74}$N. The NH-3 group hydrogen bonds with Glu$^{31}$O$^1$, the O-4 atom with Ala$^{30}$N and His$^{42}$N$^1$, and the NH$_2-5$ with Asn$^{23}$O$^1$ and His$^{42}$N$^1$. The two hydroxyl groups of the ribose have close contacts with the side chains of Asp$^{101}$ and Asn$^{103}$. The phosphate moiety forms salt bridges with His$^{76}$ and Lys$^{79}$, located in the unique insertion between the two zinc ligand cysteines, and these interactions are essential for the deamination activity because the enzyme cannot utilize the dephosphorylated form as substrate (3). These predicted substrate-binding residues are all highly conserved in the eubacterial RibGs. However, the fungal deaminases such as yeast Rib2 (yRib2) apparently contain different substrate-interacting residues because of their distinct substrate, which has an open ribityl group instead of a cyclic ribose (Figs. 1 and 5C).

Structural Divergence in the CDA Superfamily—The CDA members exist as an oligomer. All of the available member structures except for BsRibG utilize helices $\alpha$-D and surrounding loops for oligomerization. BsGD, yCD, and AaTADA display similar dimeric structures (10, 11, 14). The swapping of the C-terminal segment in BsGD causes additional dimeric contacts including the $\beta$-S strand and helices $\alpha$A and $\alpha$E. CDAs form tetramers in which one subunit interacts with the other three subunits (12). The hexameric T4dCMPD contains two types of intersubunit interfaces (13). In contrast, the D interface of BsRibG is

![Subunit interfaces. Stereo views of the D interface (A) and the R interface (B). Two distinct subunit interfaces are formed by the D and R domains with total buried areas of $\sim$1300 and $\sim$3500 Å$^2$, respectively. The D interface is made up mainly of the $\beta$-2 strand and helices $\alpha$A and $\alpha$B, whereas the R interface is made up of the two large loops, $\beta$-A and $\beta$-F, and the C-terminal tail.](image-url)
FIGURE 5. Structural conservation and divergence of the CDA superfamily. A, stereo view of structural superposition of BsRibG (red), yCD (blue) (10), T4dCMPD (green) (13), and AaTADA (yellow) (14). The zinc ion is displayed as a sphere (magenta) with the yCD inhibitor 3,4-dihydrouracil (DHU) (cyan) shown as ball-and-stick representations. These deaminases share the conserved $\beta$-sheet and helices $\alpha A$–$\alpha C$ and even a part of the $\alpha E$ helix. T4dCMPD contains a $\sim60$-residue insertion, which folds into two helices ($\alpha B'$ and $\alpha B''$) and flexible
very distinct from the others and possesses the fewest contacts, with a total buried area of ~1300 Å² (Fig. 4A). In addition to the loops and helices αA and αB, RibG includes the β2 strand in the D interface. The distinct intersubunit orientation in the D interface also separates the active sites away from each other with an inter-zinc distance of 30 Å. Notably, the shortest inter-zinc distance in the other five deaminases is about 14–15 Å.

The C-terminal segment beyond the β4 strand is quite diverse and may make a major contribution to the structural plasticity and functional diversity among the CDA members (Fig. 5A). For instance, the mononucleobase deaminase yCD and the tRNA-editing deaminase AaTADA unexpectedly superimposed very well, with an r.m.s.d. of 1.08 Å for 115 Ca atoms with 22% sequence identity. The AaTADA structure was solved by molecular replacement using yCD as a search model (14). The major structural difference around the active-site cavity is the C-terminal helix. In yCD as well as BsGD, genetic changes to alter the substrate specificity are through an introduction of substrate recognition residues at the C-terminal tail (Asp155 in yCD and Tyr156 in BsGD), which then forms a “flap,” capping and hence narrowing the opening of the active-site cavity upon substrate binding to limit the pocket size for the nucleobase (10, 11). In contrast, the C-terminal tail in AaTADA, as well as in CDAs, dCMPDs, and the D domain of BsRibG, swings away to enlarge the active-site cavity for their larger substrates (12–14). In addition to the diverse C-terminal segment, each family member has some unique substrate recognition residues; these are discussed below in detail.

**Structure-based Sequence Alignment of the CDA Members**—In combination with a sequence-structure analysis, a structure-based sequence alignment of the CDA members was constructed (Fig. 5C). Multiple sequence alignments reveal that there are unique member signatures that are useful for gene annotation. The unique substrate recognition residues for the cytosine deaminases are WXXDI at the C terminus, and those for the guanosine deaminases are FDD between helices H9251 and H9251 between helices H9251 and H9251, which then forms a “flap,” capping and hence narrowing the opening of the active-site cavity upon substrate binding to limit the pocket size for the nucleobase (10, 11). In contrast, the C-terminal tail in AaTADA, as well as in CDAs, dCMPDs, and the D domain of BsRibG, swings away to enlarge the active-site cavity for their larger substrates (12–14). In addition to the diverse C-terminal segment, each family member has some unique substrate recognition residues; these are discussed below in detail.

Even though the CDA members display substantial sequence diversity (15–25% sequence identity), our structure-sequence analysis suggests that comparative modeling is feasible. For instance, bacterial TADAs and eukaryotic TAD2s share a unique signature, EYGXVG at the N terminus of 1 strand and an extra 20–60-residue insertion between the β2 strand and the αB helix. Interestingly, most phosphate-interacting residues are located in this insertion and are not conserved across dCMPDs. RibGs have NPXVG at the N terminus of the β1 strand and an extra 8–9-residue loop between the two zinc-ligating cysteines (PCX<sub>α</sub>,C). The extra loop is highly conserved in RibGs and is predicted to interact with the phosphate moiety of the substrate.

Even though the CDA members display substantial sequence diversity (15–25% sequence identity), our structure-sequence analysis suggests that comparative modeling is feasible. For instance, bacterial TADAs and eukaryotic TAD2s share a unique signature, EXPVG, at the N terminus of the β1 strand, in which the glutamate is also conserved in TAD1. Our comparative modeling suggests that this glutamate residue may serve as an adenine recognition residue for these A-to-I tRNA deaminases.

**Possible Effects of the AID Point Mutants in Hyper-IgM Patients**—To date, apolipoprotein B mRNA-editing catalytic subunit 1 (APOBEC1) and its sequence homologues, activation-induced deaminase (AID), APOBEC2, and the tandem repeats APOBEC3A to -3H, are the only identified C-to-U RNA/DNA-editing deaminases in humans (9). AID is an essential B cell-specific factor required for antibody maturation, whereas several APOBECs are involved in defense against a broad range of retroviruses. The AID and APOBEC1 structures have been modeled with an N-terminal catalytic domain, a linker, and a C-terminal pseudo-

---

**FIGURE 6.** The NADPH-binding site in the R domain. **A**, the 2F<sub>e</sub> − F<sub>c</sub> electron density map for NADPH contoured at 1.5 σ level and is shown in cyan. **B**, schematic diagram of BsRibG interactions with the NADPH cofactor. Hydrogen bonds are presented as **dashed lines**; the interatomic distances are given in angstroms. “Radiating” **spheres** indicate hydrophobic contacts between the cofactor and the surrounding residues.

---

The C-terminal tail of yCD folds backward to limit the pocket size, whereas those of the remaining members swing away to enlarge the active-site cavity. B, superposition of the active sites of BsRibG (magenta), yCD (cyan), and T4dCMPD (green). The residue numbering is labeled in the same color for each protein. The deaminases display highly conserved interaction networks surrounding the target amino group of the nucleobase ring. In contrast, each member contains its own unique substrate recognition residues. C, multiple sequence alignment of some CDA members. The Arabidopsis thaliana deaminase (At363) (6) involved in riboflavin biosynthesis is also included. The GenBank™ accession codes are listed in the right column. Secondary structure elements for BsRibG are labeled (s). The number of residues in gaps is indicated in parentheses, and the protein length is in brackets. The superfamily signatures, H(C) and PCX<sub>α</sub>,C, are shaded in cyan, whereas the residues for the conserved hydrophobic core are in yellow. In addition, the substrate-binding residues in the known complex structures are shaded in red, and those that are predicted are in blue. The unique signatures for each member are highlighted in italics, and the residues involved in loss-of-function point mutants of AID are shaded in magenta.
Crystall Structure of RibG

FIGURE 7. Structural conservation and divergence between RibG and DHFR. A, stereo view of structural superposition of the R domain of BsRibG (red) and TmDHFR (green) (20). The R domain contains an extra ~25-residue insertion, which folds into the αD helix and the βD strand. B, multiple sequence alignment. Three other reductases involved in the riboflavin biosynthesis from A. thaliana (At599), yeast (5), and the Archaea Methanocaldococcus jannaschii (7) and two DHFRs from hyperthermophilic T. maritima and mesophilic E. coli are also included. The SwissProt accession codes are listed in the right column. Residues involved in NADPH binding are shaded in magenta, whereas the residues for the conserved hydrophobic core are in yellow. The substrate-binding residues in the known complex structures are shaded in red, and those that are predicted are in blue. DHFRs and RibG share four conserved regions for NADPH binding (motifs A–D). Their residues involved in substrate recognition are located in similar spatial positions in the βA strand, the αB helix, and the loop between the αC helix and the βC strand.

catalytic domain based on the CDA structures (30, 31). However, both proteins seem to contain the αD helix (Fig. 5C), and hence the β4 and β5 strands are parallel and the direction of the C terminus is opposite from that of CAs. In addition, both AID and APOBEC1 are too short to accommodate two CDA folds in one protein molecule. We have modeled residues 1–160 of human AID into a CDA fold consisting of the five β-strands surrounded by helices αA–αE. APOBECs contain more conserved aromatic residues than other members with a unique signature TWY[F,X]SWSPCX,C around the β3 strand.

Most residues involved in loss-of-function point mutants of AID in patients with hyper-IgM syndrome type 2 (32) are highly conserved in APOBECs (Fig. 5C). His56 and Cys87 ligate the zinc ion. Arg24, prior to the β1 strand, and Arg112 at the C terminus of the β4 strand may interact with the phosphate group of the edited cytidylate, with respect to the corresponding Ser21 and Tyr153 in T4dCMPD. Phe151 in the αE helix may form close contacts with the substrate. The conserved FFX,R motif in the αE helix in bacterial TADAs has been shown to play a critical role in A34-to-I34 deamination (33). In contrast, Trp80, Leu106, and Met139, located in the β3, β4, and β5 strands, respectively, may be responsible for construction of the conserved hydrophobic core. The truncated C147X mutant could not be expressed in 293T cells (32), implying that this mutant might not form a stable CDA fold. The 181X and 190X mutants retain deamination activity but are defective for class switch recombination and normal nuclear distribution, indicating the functional roles of the extra C-terminal segment in AID. Similarly, Ser10 and Lys10 at the N-terminal αA helix are expected to be exposed at the protein surface and might be important for nuclear location signaling or for interaction with associate proteins.

NADPH Binding in the R Domain—NADPH was identified by its strong electron density and is firmly embedded at the active site of the R domain (Fig. 6A). The enzyme did not show significant structural changes upon NADPH binding, with an r.m.s.d. of 0.4 Å for all of the backbone atoms. The cofactor is bound in an extended conformation with extensive interactions with the protein. The adenine N-3 and N-6 atoms interact with Thr221 and Ser298, respectively (Fig. 6B). The phosphate group at the ribose O-3’ makes close contacts with Thr221. These interactions seem not to contribute significantly to the cofactor binding, because NADH as well as NADPH can serve as the coenzyme (4). The pyrophosphate moiety interacts with a constellation of backbone NH groups from Gly194, Thr195, Gly292, Ser293, Ala294, and Val295 and the side chain of Thr195. These residues are located at the N termini of helices αC and αE, and the helix dipoles may partially neutralize the negative charges of the phosphate groups. The nicotiamide amide group complementarily interacts with the amide backbone of Ala153, whereas the nicotiamide ribose OH-2’ and OH-3’ contact with Asp159 and Gly165, respectively.

Structural Conservation and Divergence between BsRibG and DHFRs—A structural homology search by DALI (34) revealed that the R domain displays significant structural similarity to DHFRs (Fig. 7A). Despite low sequence identity, which is as little as 20%, all the DHFR structures exhibit a virtually identical fold consisting of a central eight-stranded β-sheet (designated βA–βH) and four flanking helices (αB, αC, αE, and αF) (28). The R domain of BsRibG contains an extra 25-residue insertion between the βD strand and the αE helix, forming the αD’ helix and the βD’ strand. This insertion is quite a distance away from the active site and is diverse across the bacterial and fungal reductases.

TmDHFR is the most stable DHFR isolated thus far. The most prominent feature of TmDHFR is its highly stable dimer, which has been shown to make a major contribution to the high intrinsic stability (20). The R-R structure of BsRibG is similar to the dimeric TmDHFR structure. Both TmDHFR and BsRibG utilize similar regions for the subunit association. The contact areas of the R interface are larger than those of TmDHFR, ~3500 Å² versus ~3050 Å², perhaps due to a larger L[4F,8G] loop in BsRibG. Structural superposition of RibG and TmDHFR reveals an r.m.s.d. of 1.5 Å for 120 Ca atoms with 25% sequence identity. The strong conservation of the tertiary structures suggests that the two reductases involved in the riboflavin and folate biosyntheses are descended from a single ancestral gene and thereby define a new superfamily.
DHFRs and BsRibG share four conserved regions for NADPH binding, motifs A–D (Fig. 7B). In motif A, the two hydrogen bonds between the amide groups of the nicotinamide ring and an alanine residue at the C terminus of the βA strand (Ala153 in BsRibG) are strictly conserved. Both proteins use motifs B and D for the pyrophosphate binding, and motif C for the phosphate group at the adenosine O-3’. The interactions between DHFRs and the 3’-phosphate group are so extensive via 3–4 residues that the enzyme has a strict NADPH dependence. Both DHFR and RibG utilize many main chain atoms for NADPH binding, and hence the corresponding sequences of the NADPH-interacting residues have diverged during evolution, but their spatial positions have remained convergent in the four regions. A conserved cis peptide bond occurs at the two consecutive glycine residues in motif D. In contrast, both enzymes form diverse interactions with the remaining areas including the adenine ring and the riboses.

The enzymatic mechanism of DHFR has been chemically and structurally studied in detail (28, 35). A complete catalytic scheme through five kinetically observable intermediates has been proposed. Similarly, RibG is expected to catalyze the reduction of a cyclic ribosyl group to an open ribityl group by hydride transfer from the C-4 atom of the nicotinamide ring of NAD[P]H to the C-1’ of the ribose with concomitant protonation of O-5’. The binding architecture of the nicotinamide ring is virtually identical in DHFR and RibG. Thus, based on the structural comparison, the substrate of the R domain was modeled into the active site with subsequent energy minimization. Simulation of the complex structure suggested that the O-2 atom of the pyrimidine ring interacts with Lys151 Nδ and the NH-3 and O-4 with Thr172 Oγ1. Thr195 and Asp199 make close contacts with the two hydroxyl groups of the ribose. Arg176, Arg183, and Arg206 form salt bridges with the phosphate group, interactions that are essential for the reduction activity because the enzyme cannot utilize the dephosphorylated form as substrate (3).

These substrate-binding residues are highly conserved in the reductases and are located in spatial positions similar to those in DHFRs. The D and R domains within the same molecule are even less interactive. Therefore, even though the two enzyme domains can fold independently, the domain fusion is crucial for the enzyme activities through formation of a stable tetrameric structure.

Acknowledgments—We thank Dr. R. Kirby for critical reading of the manuscript and Dr. H. Chang for performing the ultracentrifuge experiments.

REFERENCES

1. Bacher, A., Eberhardt, S., Fischer, M., Kie, K., and Richter, G. (2000) Annu. Rev. Nutr. 20, 153–167
2. Oltmanns, O., and Bacher, A. (1972) J. Bacteriol. 110, 818–822
3. Burrows, R. B., and Brown, G. M. (1978) J. Bacteriol. 136, 657–667
4. Richter, G., Fischer, M., Krieger, C., Eberhardt, S., Lüttingen, H., Gerstenschiöller, L., and Bacher, A. (1997) J. Bacteriol. 179, 2022–2028
5. Baur, A., Schaff-Gerstenschlager, I., Boles, E., Nüsag, T., Rose, M., and Zimmermann, F. K. (1993) Yeast 9, 289–293
6. Fisher, M., Römisch, W., Saller, S., Illarionov, B., Richter, G., Rohdich, F., Eisenreich, W., and Bacher, A. (2004) J. Biol. Chem. 279, 36292–36308
7. Graupner, M., Xu, H., and White, R. H. (2002) J. Biol. Chem. 284, 1952–1957
8. Keegan, L. P., Leroy, A., Sproul, D., and O’Connell, M. A. (2004) Genome Biol. 5, 209–218
9. Harris, R. S., and Liddament, M. T. (2004) Nat. Rev. Immunol. 4, 868–877
10. Ko, T. P., Lin, J. J., Hu, C. Y., Hsu, Y. H., Wang, A. H. J., and Liaw, S. H. (2003) J. Biol. Chem. 278, 19111–19117
11. Liaw, S. H., Chang, Y. J., Lai, C. T., Chang, H. C., and Chang, G. G. (2004) J. Biol. Chem. 279, 35479–35485
12. Johansson, E., Meijhede, N., Neuhard, J., and Larsen, S. (2002) Biochemistry 41, 2563–2570
13. Almg, M., Maley, F., Maley, G. F., Maccoll, R., and Van Roey, P. (2004) Biochemistry 43, 13715–13723
14. Kuratani, M., Ishii, R., Bessho, Y., Fukunaga, S., Sengoku, T., Shirouzu, M., Sekine, S., and Yokoyama, S. (2005) J. Biol. Chem. 280, 16002–16008
15. Kelley, L. A., MacCallum, R. M., and Sternberg, M. J. (2000) J. Mol. Biol. 299, 499–520
16. Kompsi, I. M., Islam, G., and Then, R. L. (2005) Chem. Rev. 105, 593–620
17. Schuck, P., Perugini, M. A., Gonzales, N. R., Howlett, G. J., and Schubert, D. (2002) Biophys. J. 82, 1096–1111
18. Tervilliger, T. C. (2003) Methods Enzymol. 34, 22–37
19. Wang, J. W., Chen, J. R., Gu, Y. X., Zheng, C. D., Jiang, F., Fan, H. F., Tervilliger, T. C., and Hao, Q. (2004) Acta Crystallogr. Sect. D 60, 1244–1253
20. Dams, T., Auerbach, G., Bader, G., Jacob, U., Ploom, T., Huber, R., and Jaenicke, R. (2000) J. Mol. Biol. 297, 659–672
21. Roussel, A., and Cambillau, C. (1991) in Silicon Graphics Geometry Partners Directory RI, Silicon Graphics Corp., Mountain View, CA
22. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Acta Crystallogr. Sect. D 54, 905–921
23. Schäfer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. I., Koonin, E. V., and Altschul, S. F. (2001) Nucleic Acids Res. 29, 2995–3005
24. Chenna, R., Sugawara, H., Koike, T., Lopes, R., Gibson, T. J., Higgins, D. G., and Thompson, J. D. (2003) Nucleic Acids Res. 31, 3497–3500
25. Enosuf, R. M. (1999) Acta Crystallogr. Sect. D 55, 938–940
26. Merritt, E. A., and Bacon, D. J. (1997) Methods Enzymol. 277, 505–524
27. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) Protein Eng. 8, 127–134
28. Schnell, J. R., Dyson, H. J., and Wright, P. E. (2004) Annu. Rev. Biophys. Biomol. Struct. 33, 119–140
29. Wójtan, D. W., Greasy, S. E., Beardsley, G. P., and Wilson, I. A. (2002) J. Bacteriol. 145, 1505–1515
30. Navaratnam, N., Fujino, T., Bayliss, J., Jarmuz, A., Bessho, Y., Fukunaga, S., Sengoku, T., Shirouzu, M., Sekine, S., and Yokoyama, S. (2005) J. Biol. Chem. 280, 20222–20228
31. Xu, H., Svarovskaia, E. S., Barr, R., Zhang, Y., Khan, M. A., Strebel, K., and Pathak, V. K. (2004) Proc. Natl. Acad. Sci. U. S. A. 101, 8562–8567
32. Ta, Y. T., Nagaoaka, H., Catalan, N., Durandy, A., Fischer, A., Imam, K., Nonoyama, S., Tashiro, J., Iegawa, M., Ito, S., Kinoshita, K., Muramatsu, M., and Honjo, T. (2004) Nat. Immunol. 5, 843–848
33. Elias, Y., and Huang, R. H. (2005) Biochemistry 44, 12057–12065
34. Dietmann, S., Park, J., Notredame, C., Heger, L., Marme, F., and Holm, L. (2001) Nucleic Acids Res. 29, 55–57
35. Fierke, C. A., Johnson, K. A., and Benkovic, S. J. (1987) Biochemistry 26, 4085–4092
36. Yuvarajyam, J., Chitnumsub, P., Kamchonwongpaisan, S., Vanichthanakul, J., Sirawaraporn, W., Taylor, P., Walkinshaw, D. M., and Yathavong, Y. (2003) Nat. Struct. Biol. 10, 357–365

Crystal Structure of RibG

MARCH 17, 2006•VOLUME 281•NUMBER 11 JOURNAL OF BIOLOGICAL CHEMISTRY 7613