Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Evolution of vitamin B_2 biosynthesis: eubacterial RibG and fungal Rib2 deaminases

Sheng-Chia Chen,[a]‡ Chieh-Yi Shen,[a]‡ Te-Ming Yen,[b] Hui-Chia Yu,[a] Ting-Hao Chang,[a] Wen-Lin Lai[c,d]* and Shwu-Huey Liaw[a,e]*

[a]Department of Life Sciences and Institute of Genome Sciences, National Yang-Ming University, Taipei 11221, Taiwan, [b]Institute of Biochemistry and Molecular Biology, National Yang-Ming University, Taipei 11221, Taiwan, [c]Clinical Laboratory, Chung-Shan Medical University Hospital, Taichung 40201, Taiwan, [d]School of Medical Laboratory and Biotechnology, Chung-Shan Medical University, Taichung 40201, Taiwan, and [e]Department of Medical Research and Education, Taipei Veterans General Hospital, Taipei 11217, Taiwan

‡ These authors contributed equally to this work.

Correspondence e-mail: wllai@csmu.edu.tw, shliaw@ym.edu.tw

Received 18 July 2012
Accepted 30 October 2012

PDB Reference: RibG, 4g3m

Eubacterial RibG and yeast Rib2 possess a deaminase domain for pyrimidine deamination in the second and third steps, respectively, of riboflavin biosynthesis. These enzymes are specific for ribose and ribitol, respectively. Here, the crystal structure of *Bacillus subtilis* RibG in complex with a deaminase product is reported at 2.56 Å resolution. Two loops move towards the product on substrate binding, resulting in interactions with the ribosyl and phosphate groups and significant conformational changes. The product carbonyl moiety is bent out of the pyrimidine ring to coordinate to the catalytic zinc ion. Such distortions in the bound substrate and product may play an essential role in enzyme catalysis. The yeast Rib2 structure was modelled and a mutational analysis was carried out in order to understand the mechanism of substrate recognition in these two enzymes. Detailed structural comparisons revealed that the two consecutive carbonyl backbones that occur prior to the PC*XX*C signature constitute a binding hole for the target amino group of the substrate. This amino-binding hole is essential in *B. subtilis* RibG and is also conserved in the RNA/DNA-editing deaminases.

1. Introduction

Riboflavin serves as the direct precursor of the FAD/FMN cofactors that are involved in a wide variety of essential cellular processes. This vitamin (B_2) is biosynthesized in most prokaryotes, fungi and plants (Bacher *et al.*, 2000; Fischer & Bacher, 2006). Some pathogenic microorganisms lack an efficient uptake system and can only obtain this vitamin by biosynthesis. In contrast, animals lack this biosynthetic pathway and hence depend on riboflavin uptake from dietary sources. Therefore, the enzymes that are involved in riboflavin biosynthesis appear to be attractive targets for antimicrobial drug design (Mack & Grill, 2006; Kaiser *et al.*, 2007; Long *et al.*, 2010) and thus have been the subject of considerable investigation over the years.

Most commercial riboflavin is currently produced by special strains of *Bacillus subtilis*, *Ashbya gossypii* and *Candida famata* (Abbas & Sibirny, 2011). The biosynthetic pathways are identical in plants and eubacteria, but differ in archaea and fungi. The first committed step is catalyzed by GTP cyclohydrolase II (GCH II) to yield formate, pyrophosphate and 2,5-diamino-6-ribosylamino-4(3*H*)-pyrimidinone 5′-phosphate (DAROPP; Richter *et al.*, 1993). Most eubacteria possess a bifunctional protein such as *B. subtilis* RibG (BsRibG; Richter *et al.*, 1997) with an N-terminal deaminase domain for the

deamination of DAROPP to 5-amino-6-ribosylamino-2,4(1*H*,3*H*)-pyrimidinedione 5′-phosphate (AROPP) and a C-terminal reductase domain for subsequent NAD(P)H-dependent reduction to 5-amino-6-ribitylamino-2,4(1*H*,3*H*)-pyrimidinedione 5′-phosphate (ARIPP) (Fig. 1). In contrast, in fungi and some archaea DAROPP is first reduced to 2,5-diamino-6-ribitylamino-4(3*H*)-pyrimidinone 5′-phosphate (DARIPP) and then deaminated to ARIPP. The enzymes responsible for these reactions are two separate proteins: Rib7 and Rib2 (Graupner *et al.*, 2002; Behm-Ansmant *et al.*, 2004). Yeast Rib2 (yRib2) is also a bifunctional protein, with an N-terminal pseudouridine synthase domain and a C-terminal deaminase domain.

Recently, we have solved the tetrameric structure of BsRibG with substrate bound in the reductase domain (Chen *et al.*, 2009). The reductase domain and the pharmaceutically important enzyme dihydrofolate reductase share a well conserved core structure, NADPH-binding site, catalytic mechanism and even substrate-binding architecture. On the other hand, the deaminase domain belongs to the cytidine deaminase (CDA) superfamily, which includes several mononucleotide deaminases involved in nucleotide metabolism and RNA/DNA-editing deaminases associated with gene diversity and antivirus defence (Conticello, 2008). RNA/DNA-editing deaminases are comprised of A-to-I tRNA-specific adenosine deaminases (TADs), A-to-I adenosine deaminases acting on RNA (ADARs) and the C-to-U deaminases of the apolipoprotein B mRNA-editing catalytic protein (APOBEC) family (Nishikura, 2010; Smith *et al.*, 2012). These enzymes catalyze zinc-dependent hydrolytic deamination of the base moiety of a variety of nucleotides. The consensus histidine and cysteine residues in the H(C)*X*E and PC*XX*C superfamily signatures coordinate the zinc ion, while the glutamate assists in proton transfer.

To gain structural insight into the distinct substrate specificity of eubacterial RibG and fungal Rib2, we determined the crystal structure of BsRibG with the product bound to the deaminase domain at 2.56 Å resolution. We then modelled the yRib2 deaminase and carried out a mutational analysis. In addition, detailed structural comparisons suggested the presence of a conserved amino-binding hole in the CDA superfamily.

Table 1
Statistics of data collection and structural refinement.

Values in parentheses are for the highest resolution shell.

PDB code	4g3m
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 88.0, b = 109.2, c = 189.5
Resolution (Å)	50–2.56 (2.65–2.56)
Observed reflections	395937 (37245)
Unique reflections	58717 (5730)
Completeness (%)	99.2 (97.9)
Data collection	
$\langle I/\sigma(I)\rangle$	30.7 (3.2)
R_{merge}† (%)	5.4 (59.9)
Refinement	
Resolution (Å)	50–2.56 (2.65–2.56)
Reflections ($F > 0\sigma(F)$]	56431 (4194)
R_{cryst}‡ (%)	22.2 (31.9)
R_{free}§ (%)	27.0 (36.8)
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.2
Mean B values (Å^2)	
Protein atoms (11009)	57.9
Zinc ions (4)	56.5
AROPP	70.9
ARIPP intermediate	71.8
Water molecules (345)	56.0
Ramachandran analysis¶ (%)	
Favoured	93.1
Allowed	6.1
Disallowed	0.8

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the average intensity value of the equivalent reflections. ‡ $R_{cryst} = \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where F_{obs} and F_{calc} are observed and calculated structure factors, respectively. § R_{free} was calculated using 10% of data randomly excluded from refinement. ¶ The Ramachandran analysis was performed using *MolProbity* (Chen *et al.*, 2010).

2. Materials and methods

2.1. Preparation of protein and substrates

Site-directed mutagenesis was carried out using a QuikChange site-directed mutagenesis kit (Stratagene). The recombinant wild-type and mutant BsRibG and yRib2, *Methanosarcina mazei* Rib7 and *Escherichia coli* GCH II were expressed in *E. coli* using the pQE-30 vector (Qiagen). Luria broth cultures were grown at 310 K and induced with 1 m*M* isopropyl β-D-1-thiogalactopyranoside. Cell pellets were resuspended in lysis buffer consisting of 10 m*M* HEPES, 100 m*M* NaCl pH 7.5 and lysed using a French press. The crude extract was applied onto an Ni–NTA (Qiagen) column. After washing with 20–60 m*M* imidazole,

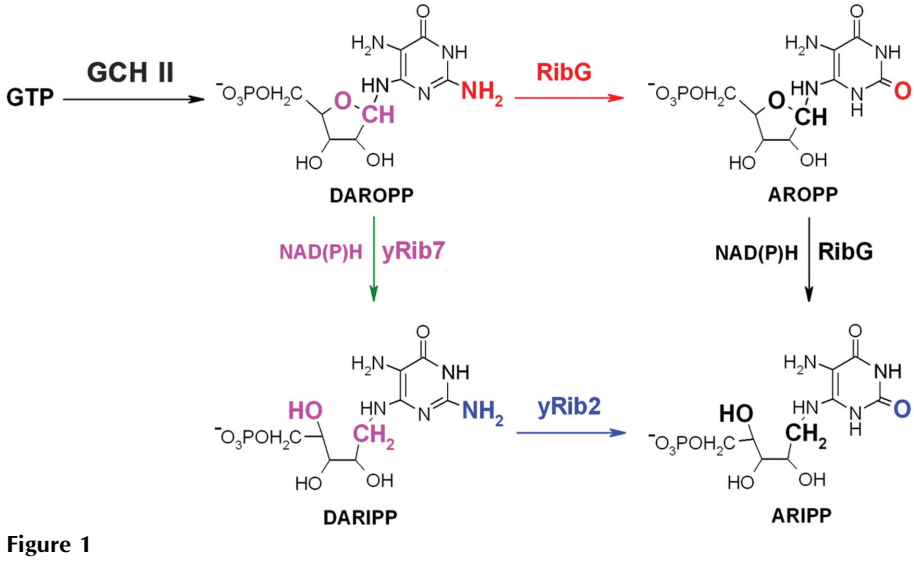


Figure 1
Pyrimidine deamination and ribosyl reduction in vitamin B_2 biosynthesis.

the protein was eluted with 300 m*M* imidazole and dialyzed against the lysis buffer, with the exception of *E. coli* GCH II, for which 10 m*M* $MgCl_2$, 100 m*M* Tris–HCl pH 8.5 was used. An additional Q-Sepharose anion-exchange column (GE Healthcare) using an NaCl gradient was applied to improve protein purity when necessary.

DAROPP was prepared by the addition of 0.5 *M* GTP to a final concentration of 10–30 m*M* to *E. coli* GCH II solution (5 mg ml^{-1}). The reaction was complete in about 10–15 min; recombinant (41 kDa) BsRibG (2 mg) was then added for 10–15 min to produce AROPP, while *M. mazei* Rib7 (1 mg) and 25 m*M* NADPH were added for 5 min to produce DARIPP. GTP, DAROPP and AROPP display distinct absorption spectra with absorption maxima of 254, 292 and 284 nm, respectively, and hence the the compound preparation was monitored by UV–Vis absorption spectra recorded on a Beckman DU640B spectrophotometer. The concentration of DAROPP was estimated using its absorption coefficient of 9600 M^{-1} cm^{-1} at 292 nm, while the concentrations of AROPP and DARIPP were estimated from the consumption of NADPH catalyzed by the BsRibG reductase domain and by *M. mazei* Rib7, respectively. After the reaction, the enzymes were removed using a vial filter with a 10 kDa cutoff membrane and 10–30 m*M* dithiothreitol was added to prevent oxidation.

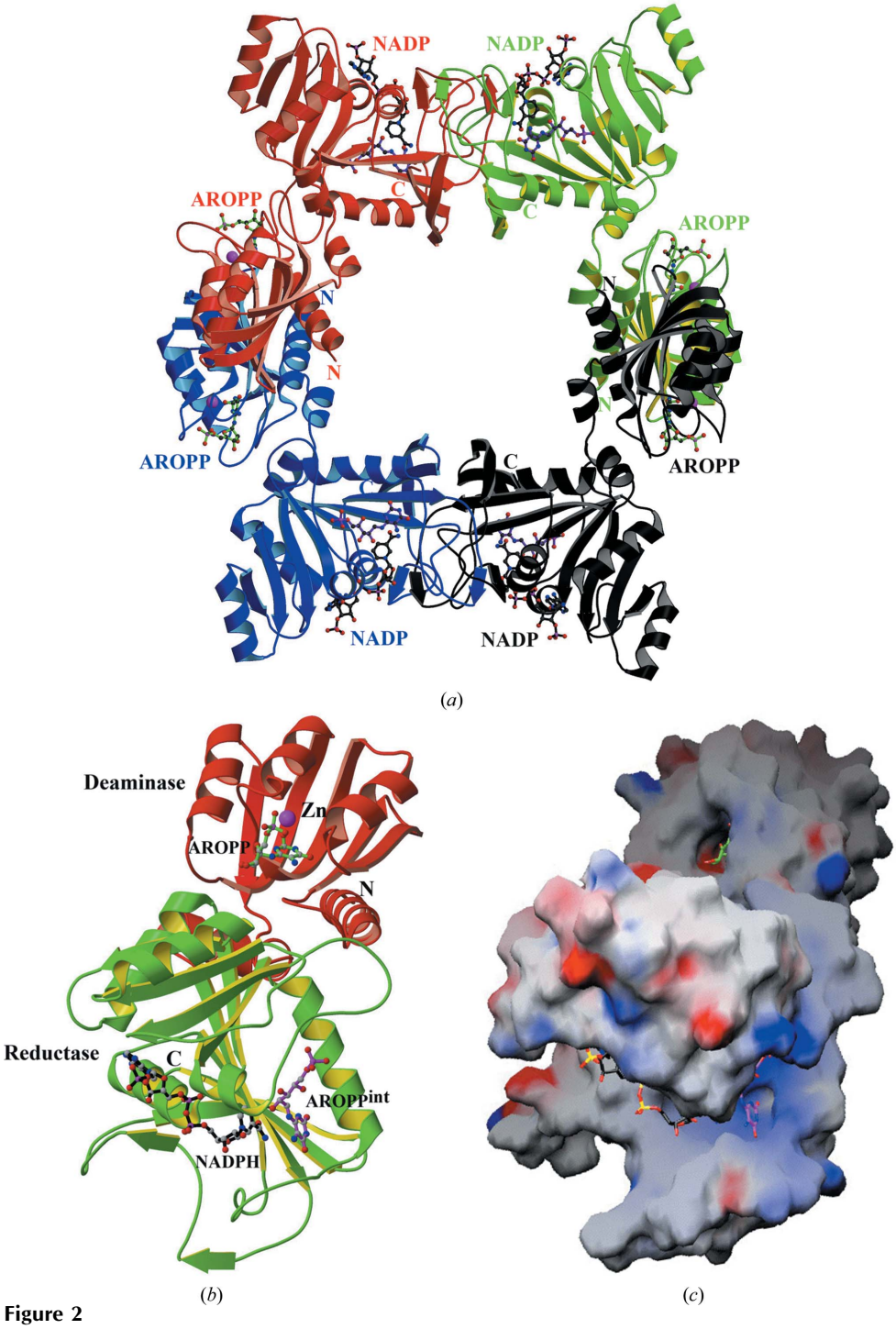


Figure 2
BsRibG tetramer (*a*) and monomer (*b*) structures. (*c*) Molecular surface of one monomer. The view is the same as that in (*b*), with electrostatic potential from $-20k_BT$ (red) to $20k_BT$ (blue). Together with the previous NADPH binary structure, a ternary structure with AROPP (green) and a zinc ion (magenta) bound in the deaminase domain and with the ribitylimino intermediate (magenta) and the cofactor NADPH (black) bound in the reductase domain was generated.

2.2. Activity assay

The relative deaminase activity of wild-type and mutant BsRibG was measured by two methods. In the first method, various amounts of BsRibG and substrate were incubated in a solution consisting of 0.1 *M* Tris–HCl pH 7.5, 0.1 *M* NaCl, 10 m*M* dithiothreitol, and the decrease in A_{340} resulting from NAD(P)H oxidation owing to the overall reaction of the deaminase-reductase was monitored. The other assay was a chromatographic analysis using a Cosmosil 5C_{18}-AR-300 column (4.6 × 250 mm) equipped with an ÄKTApurifier (GE Healthcare). Virtually identical protocols were used as described previously (Magalhães *et al.*, 2008), with a flow rate of 0.8 ml min^{-1}. GTP, GMP, DAROPP, AROPP and ARIPP eluted from the column at retention volumes of 23.2, 12.8, 9.4, 11.5 and 11.8 ml, respectively.

DARIPP eluted as a double peak at 10–11 ml.

The activities of wild-type and mutant yRib2 were also assayed using two methods. C_{18} column analysis was used for the first analysis, while in the second method changes in fluorescence emission after diacetyl modification were measured based on the observation that diacetyl derivatives of DAROPP, DARIPP, AROPP and ARIPP display distinct fluorescence spectra with various wavelength maxima for excitation and emission (Richter *et al.*, 1997). The reaction mixtures were incubated as described above for 30–90 min at 310 K. Diacetyl was added to a final concentration of 1%(*v*/*v*) and the mixture was incubated at 368 K for 30 min. The fluorescence spectra were recorded using a Hitachi F-4500 spectrophotometer.

2.3. Structure analysis

The BsRibG crystals were first grown in 28% polyethylene glycol 400, 200 m*M* $MgCl_2$, 5% glycerol, 100 m*M* HEPES pH 7.5 using a protein concentration of 20–25 mg ml^{-1} (Chen *et al.*, 2006). The AROPP derivative was prepared by soaking crystals for 1–2 h in reservoir solution containing 10–15 m*M* AROPP. X-ray diffraction data were collected and processed on beamlines BL13B1 and BL13C1 at NSRRC, Hsinchu, Taiwan. The structure was determined by molecular replacement using the apo BsRibG structure (PDB entry 2b3z; Chen *et al.*, 2006) as the initial model and was refined using *CNS* (Brunger, 2007). The statistics of data collection and refinement are summarized in Table 1. Figs. 2(*a*), 2(*b*), 3(*b*), 4(*a*), 6(*b*), 7, 8(*a*) and 8(*b*) were generated using *MolScript* (Kraulis, 1991) and *Raster3D* (Merritt & Bacon, 1997), Fig. 2(*c*) using *GRASP* (Nicholls *et al.*, 1991), Fig. 3(*a*) using *BobScript* (Esnouf, 1997) and Fig. 4(*b*) using *PyMOL* (DeLano, 2002).

3. Results

3.1. Product binding to the deaminase domain

BsRibG exists as a tetramer in solution as well as in the crystal, in which the four subunits in the asymmetric unit form a ring-like tetramer (Chen *et al.*, 2006; Fig. 2). Each protomer is composed of an N-terminal deaminase domain and a C-terminal reductase domain. The soaking of native BsRibG crystals with AROPP and subsequent structure determination and calculation of the electron-density maps revealed a blob-like density which could be accounted for by the product AROPP and a ribitylimino intermediate in the deaminase and reductase domains, respectively. The interaction network between the ribitylimino intermediate and the reductase domain is virtually identical to that observed in our previously reported structure (Chen *et al.*, 2009).

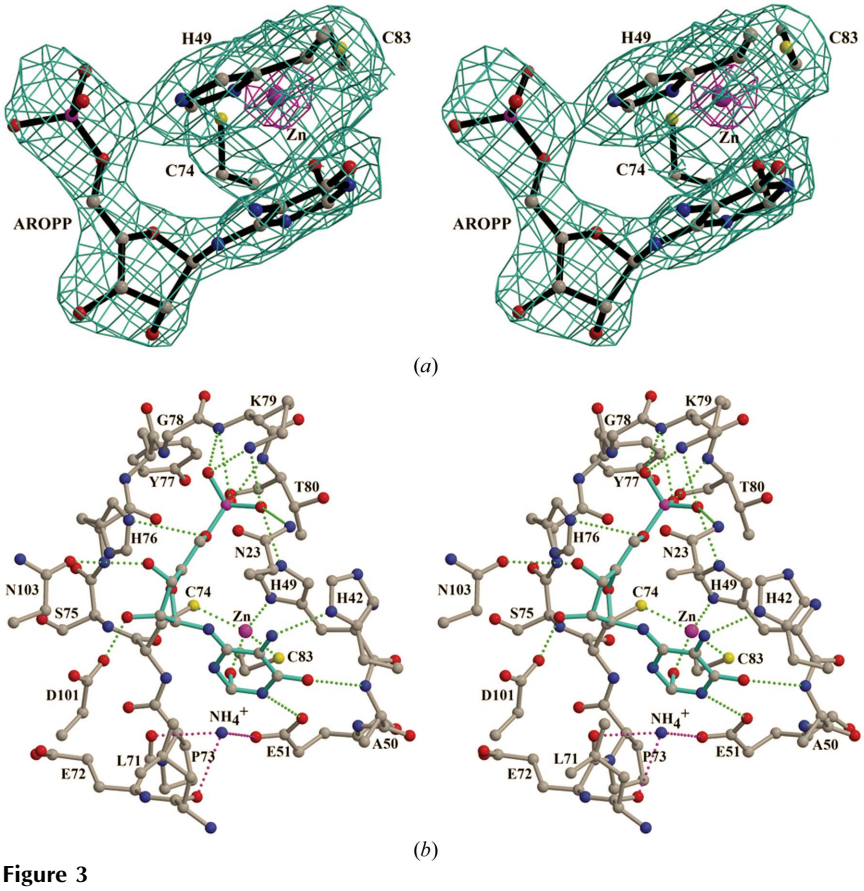


Figure 3
The AROPP binding site in the deaminase domain of BsRibG. (*a*) Stereoview of the simulated-annealing $2F_o - F_c$ OMIT map of the active site contoured at 1.0σ and 6σ levels and shown in cyan and magenta, respectively. (*b*) Interaction networks between AROPP and the deaminase domain in protomer *B*. Hydrogen bonds are shown as green dashed lines. The proposed product ammonium ion is displayed as a blue sphere and its potential interactions with Glu51 O$^{\varepsilon1}$, Leu71 O and Glu72 O are highlighted by magenta dashed lines.

In the deaminase domain, AROPP could be modelled nicely into the electron density in protomers *B*–*D* (Fig. 3*a*), whereas no such modelling was possible in protomer *A*. The tightly bound endogenous zinc ion is coordinated by His49 N$^{\delta1}$ (2.2 Å), Cys74 S$^{\gamma}$ (2.4 Å), Cys83 S$^{\gamma}$ (2.2 Å) and the carbonyl moiety of the product (2.2 Å) (Fig. 3*b*). The pyrimidine N^{3}H group interacts with Glu51 O$^{\varepsilon2}$ (2.9 Å), the O^4 atom interacts with Ala50 N (3.1 Å) and the N^{5}H$_2$ group interacts with His42 N$^{\delta1}$ (3.4 Å). The O$^{2'}$H group of the ribose has close contacts with Asp101 O$^{\delta1}$ (2.5 Å) and the OH$^{3'}$ group has close contacts with Asn103 O$^{\delta1}$ (3.0 Å). The phosphate moiety forms hydrogen bonds to Asn23 N$^{\delta2}$ (3.5 Å), His49 N$^{\varepsilon2}$ (2.7 Å), His76 N$^{\varepsilon2}$ (3.1 Å), Lys79 N (3.0 Å), Lys79 N$^{\zeta}$ (2.9 and 3.2 Å) and Thr80 N (2.7 Å). In addition, the imidazole ring of His49 stacks on the pyrimidine ring with an interplanar distance of ∼3.5 Å. These AROPP-interacting residues are all conserved in the eubacterial RibGs.

Consistent with the previous NADPH binary structure (Chen *et al.*, 2006), a ternary structure with AROPP bound in

the deaminase domain and the ribitylimino intermediate and the cofactor NADPH bound in the reductase domain was generated (Fig. 2). Kinetic measurements of the *E. coli* homologue RibD showed that reduction occurs much more slowly than deamination (Magalhães *et al.*, 2008). This suggests that there is no substrate-transportation channel between the two enzyme active sites. Our BsRibG structures are consistent with the kinetic studies. The openings of the active sites of the deaminase and reductase domains point in different directions and the phosphate groups of the two substrates are ∼30 Å apart. Unlike the bifunctional enzyme encoding dihydrofolate reductase and thymidylate synthase activity in *Plasmodium falciparum* (Yuvaniyama *et al.*, 2003), this BsRibG structure does not display a channel for substrate transportation from the deaminase domain to the reductase domain that would allow sequential reactions (Fig. 2*c*).

3.2. Multiple conformations of the deaminase domain

The average *B* factors of AROPP in protomers *B*, *C* and *D* were 67, 71.8 and 74 Å^2, respectively. This suggests that the product AROPP binds in the four deaminase domains of the BsRibG tetramer with variable occupancies. Superposition of the four deaminase domains showed multiple conformations involving two loops (Fig. 4). In protomers *B*–*D* the two loops move towards AROPP, forming direct interactions, and hence significant conformational changes are induced. One of the changes involves the RibG-unique loop between the two zinc-ligand cysteine residues, which contains His76, Lys79 and Thr80 that interact with the AROPP phosphate group; therefore, we have named this the P loop. The other loop with Asp101 and Asn103 is responsible for ribosyl binding and is called the R loop. In protomer *A*, owing to crystal packing, the P loop interacts with the reductase interfaces of the adjacent subunits; the distance between Lys79 and His321 in an adjacent subunit is 4.0 Å. Hence, it is unable to move towards the substrate-binding site.

3.3. Activity assay of BsRibG deaminase mutants

The BsRibG–AROPP structure revealed that Asn23, His42, Glu51, His76, Lys79, Asp101 and Asn103 are involved in substrate binding; hence, mutants of BsRibG were constructed by individually substituting these residues with alanine. In our previous report, the relative deaminase activity was measured by monitoring the change in absorbance at 292 nm, which is the wavelength of maximum absorbance of AROPP. The results indicated that no detectable activity was present in the E51A and K79A mutants, whereas the other five mutants showed an activity that was decreased by twofold to threefold compared with the wild type (Chen *et al.*, 2009). However, the absorbance changes at 292 nm may be too small (∼0.01 for the wild type) and this might lead to large errors.

Here, we employed two assays that were more appropriate. Using AROPP as a substrate, similar reductase activities as represented by the decrease in A_{340} were detected for the seven mutants compared with the wild type. This indicates that all of the mutants possess an active reductase domain. However, even using the high protein concentration of 1.0 mg ml^{-1} and 2.5 m*M* DAROPP and 0.25 m*M* NADPH for

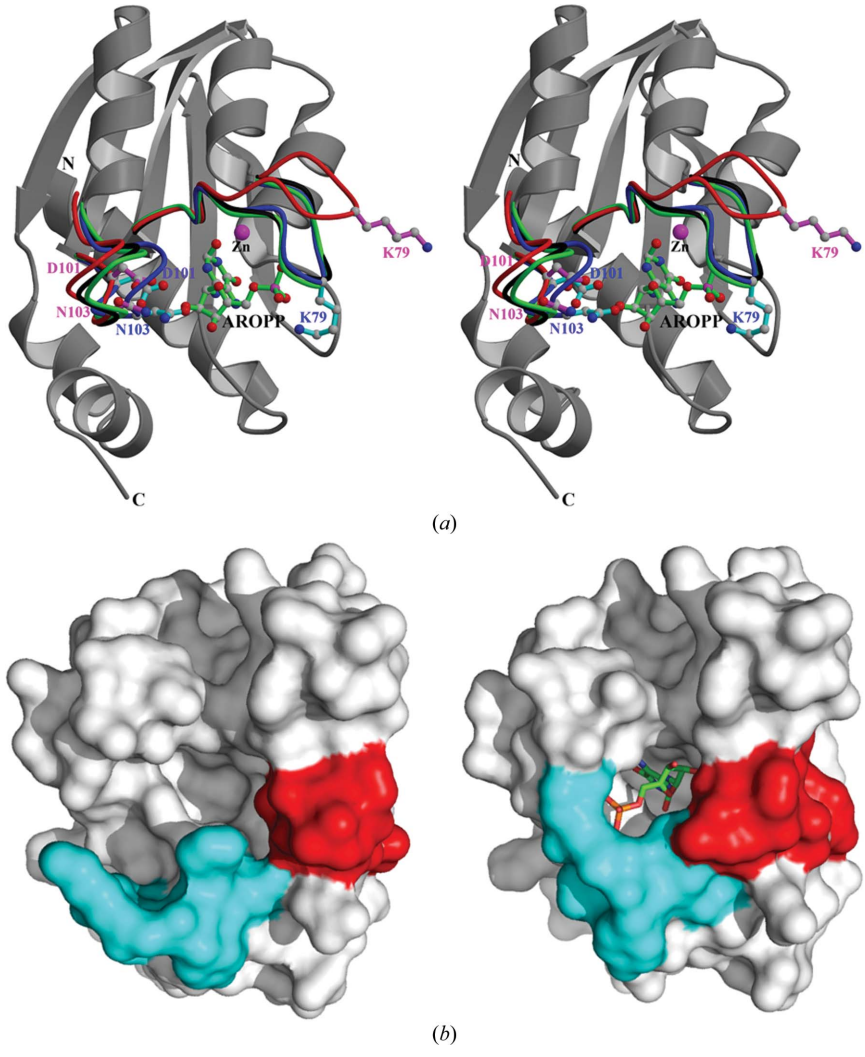


Figure 4
Multiple conformations in the deaminase domain of BsRibG. (*a*) Structural superposition of the deaminase domains in the tetramer. Significant structural differences are observable in the P and R loops. The P loop in protomer *A* (red) is different from the others because of its interaction with the adjacent subunit. On the other hand, only the R loop in protomer *B* (blue) moves towards AROPP for binding. (*b*) Molecular surfaces of the deaminase domain in molecules *A* (left) and *B* (right). The P and R loops are shown in cyan and red, respectively.

20 min, no detectable activity was observed for the N23A, H42A, E51A, H76A and D101A mutants, 0.5–1% of the activity was observed for the N103A variant and ∼2–3% of the activity was observed for the K79A variant. These results were consistent with the C_{18} column analysis. For the K79A and N103A mutants, a small amount of the product AROPP could be detected with 1.0 mg ml^{-1} protein and 2.5 m*M* DAROPP (Fig. 5). However, for the other five mutants neither detectable disappearance of DAROPP nor AROPP production was observed. Therefore, we conclude that Glu51, Asn23, His42, His76 and Asp101 are essential residues in BsRibG, while Lys79 and Asn103 are important for DAROPP deamination.

3.4. Distinct substrate specificity of yRib2

Our activity assay showed that BsRibG recognizes only DAROPP and not DARIPP as the deaminase substrate, whereas yRib2 discriminates DARIPP from DAROPP. The deaminase domains of BsRibG and yRib2 possess 145 and 158 residues, respectively, and share 32 identical residues (∼20% sequence identity; Fig. 6*a*). Based on the BsRibG–AROPP structure, we modelled the structure of yRib2 complexed with DARIPP using the *SWISS-MODEL* server (Arnold *et al.*, 2006) with molecular-dynamics simulations using the *GROMACS* suite (Bjelkmar *et al.*, 2010). Structural superposition of BsRibG and yRib2 revealed a root-mean-square deviation of 0.9 Å for 138 C^{α} atoms. They share superimposable strands and helices but have diverse loops, particularly a longer loop between the αB helix and the β3 strand in yRib2 (Fig. 6*a*). Obviously, BsRibG and yRib2 possess different substrate-binding residues for their distinct substrates (Figs. 1 and 6). However, a large portion of these residues may be located at equivalent spatial positions. Similar to the BsRibG–AROPP complex, Glu484 $O^{\varepsilon1}$, Ala483 N and Ser474 O^{γ} form hydrogen bonds to the pyrimidine N^3H, O^4 and N^5H_2 moieties, respectively, while Thr454 $O^{\gamma1}$ and Ser520 O^{γ} interact with the phosphate group (Fig. 6*b*). Multiple sequence alignment of yRib2 and other homologues with 30–60% sequence identity shows that these predicted DARIPP-interacting residues are highly conserved.

The observed multiple conformations of the P and R loops in BsRibG (Fig. 4) may imply that these two regions are flexible in RibG and its homologue Rib2 and can adopt the different conformations required for their distinct substrate specificity for cyclic ribose *versus* linear ribitol (Fig. 1). In yRib2, the conserved Glu545 and the less conserved Asp547 in the R loop, together with the invariant Arg518 in the P loop, may be involved in binding to the substrate DARIPP. The effects of the substitution of Arg518, Glu545 and Asp547 by alanine were examined. In the diacetyl assay, when the wild-type protein concentration increases both the fluorescence intensity and the shift in the wavelength increase (Fig. 6*c*), showing a two-state-like curve (Fig. 6*d*). The protein concentration at the midpoint of the transition is ∼1 µg ml^{-1} for the wild type and D547A mutant, while it is ∼100 µg ml^{-1} for the E545A variant. However, even with 1.0 mg ml^{-1} protein and ∼2.5 m*M* DARIPP the fluorescence spectrum did not display any wavelength shift for the R518A mutant. In the C_{18} column analysis, a small amount of the product ARIPP could be detected with the E545A mutant, whereas there was neither substrate consumption nor appearance of product for the R518A mutant. Therefore, Arg518 is essential in yRib2, while Glu545 is important for DARIPP deamination. For Arg518 and Glu545 binding to the DARIPP phosphate and ribityl groups, the P and R loops in yRib2 must take up different conformations from those in BsRibG. Our modelled yRib2–DARIPP structure suggested that Arg518 $N^{\eta1}$ and $N^{\eta2}$ interact with the phosphate, while Glu545 $O^{\varepsilon1}$ interacts with the ribityl $O^{2'}H$ group (Fig. 6*b*).

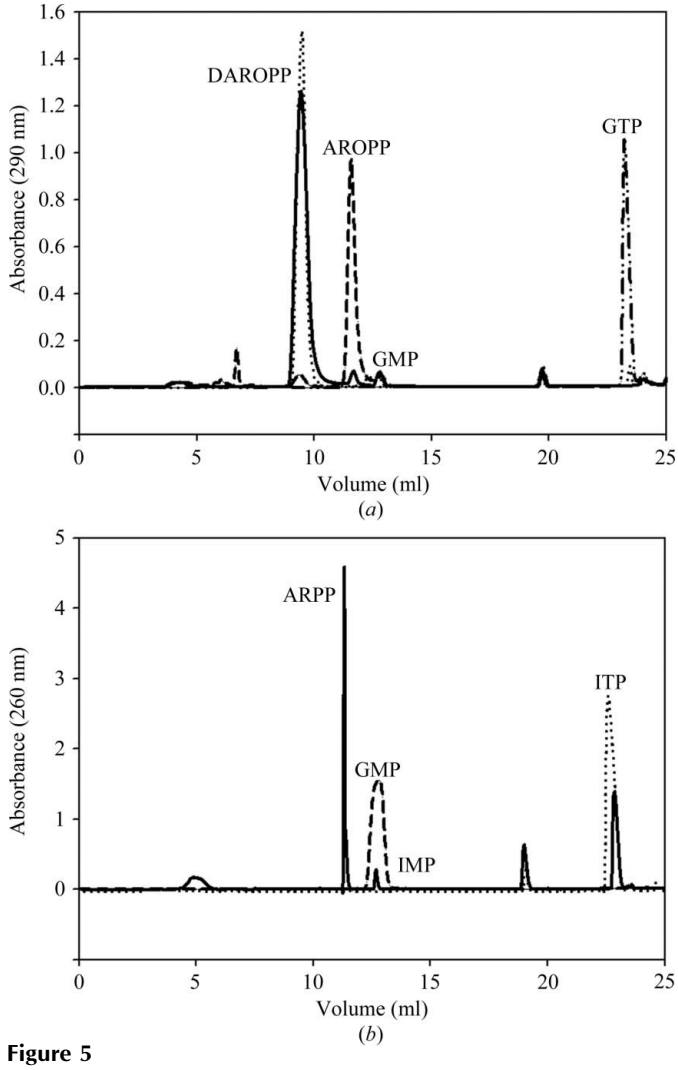
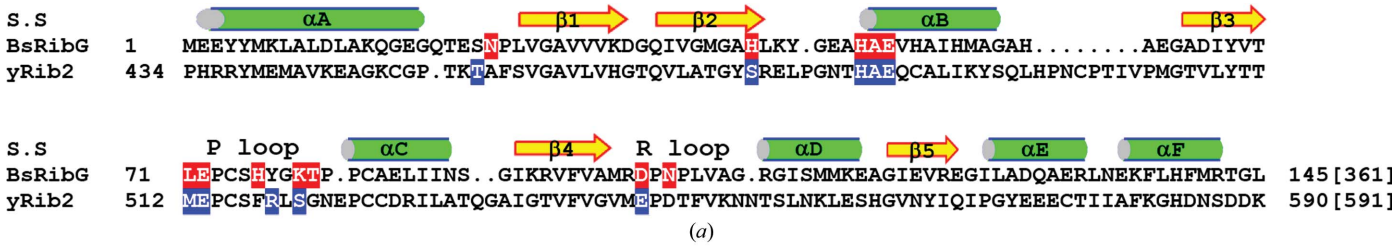
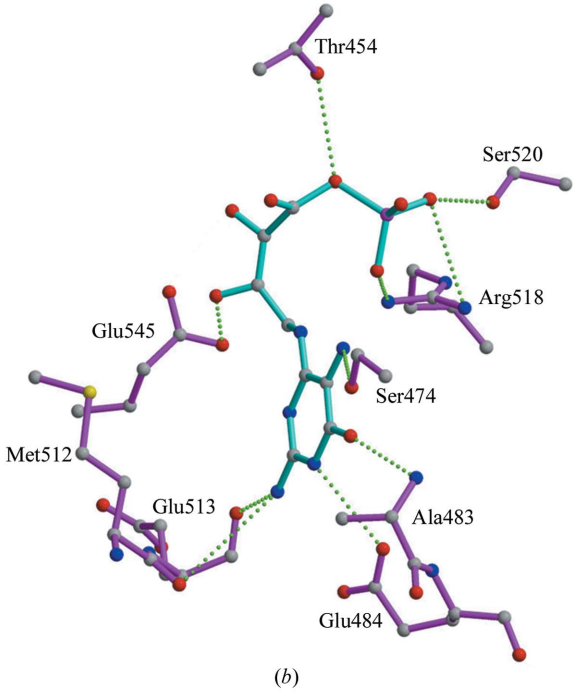


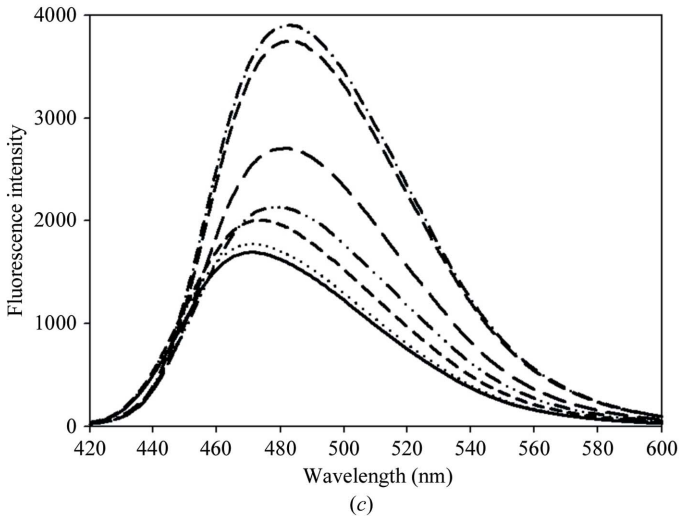
Figure 5
Activity assay by chromatographic analysis. (*a*) Deaminase assay of BsRibG. After the addition of GTP (dotted/dashed line) to the *E. coli* GCH II solution (5 mg ml^{-1}), the appearance of DAROPP and GMP was observed (dotted line). Addition of wild-type BsRibG (0.1 mg ml^{-1}) leads to the disappearance of DAROPP and the production of AROPP (dashed line). At a protein concentration of 0.1 mg ml^{-1} with 2.5 m*M* DAROPP for 30 min at 310 K, the K79A variant displayed low activity (dashed line). (*b*) The BsRibG assay with ARPP as substrate. After incubation of 1 m*M* ITP (dotted line) with *E. coli* GCH II (2 mg ml^{-1}) for 30 min at 310 K, the appearance of ARPP and IMP was observed (dashed line). However, addition of wild-type BsRibG (0.1 mg ml^{-1}) did not result in the disappearance of ARPP. Similarly, incubation of wild-type BsRibG with 2.5 m*M* GMP did not consume any GMP (dashed line).



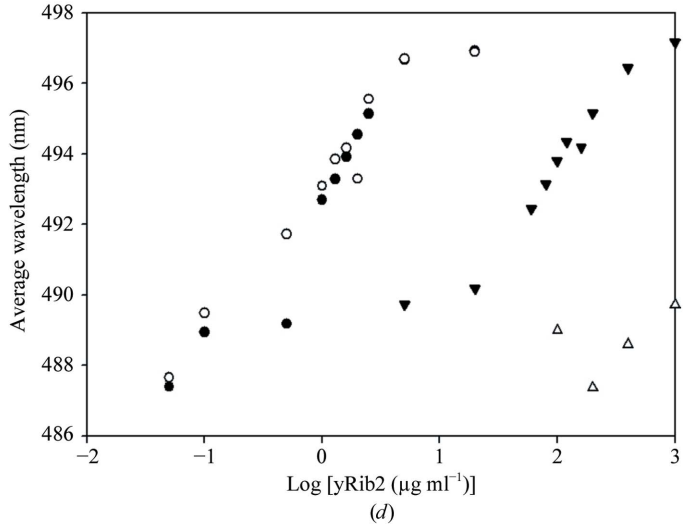
(*a*)



(*b*)



(*c*)



(*d*)

Figure 6
The predicted substrate-binding residues in yRib2. (*a*) Sequence alignment of the deaminase domains of BsRibG and yRib2. The substrate-binding residues in BsRibG are shown in red, while those predicted in yRib2 are shown in blue. These predicted residues are highly conserved in fungal deaminases and are located in similar regions as those in BsRibG. (*b*) The predicted interaction networks between DARIPP and yRib2. (*c*) The fluorescence spectra after the diacetyl modification with an excitation wavelength of 408 nm. The spectra were recorded with 0.5 m*M* DARIPP and various wild-type yRib2 concentrations: control (unbroken line), 0.05 μg ml^{-1} (dotted line), 0.5 μg ml^{-1} (short-dashed line), 1.3 μg ml^{-1} (dash/dot/dotted line), 2.5 μg/ml (medium-dashed line), 5 μg ml^{-1} (long-dashed line) and 20 μg ml^{-1} (dash/dotted line). (*d*) A plot of the average fluorescence emission wavelength against protein concentration: wild type (filled circles), D547A (circles), E545A (filled inverted triangles) and R518A (triangles). A log scale was used because the protein concentrations spanned several orders of magnitude.

3.5. An essential amino-binding hole in BsRibG

On the basis of the BsRibG–AROPP and mouse CDA–cytidine structures (Teh *et al.*, 2006), the substrate DAROPP and the transition-state intermediate were modelled into the active site. Subsequent molecular-dynamics simulations were performed using the *GROMACS* suite (Bjelkmar *et al.*, 2010). The target amino groups interact with two consecutive backbone carbonyl groups prior to the consensus PC*XX*C motif (Leu71 and Glu72 in BsRibG; Figs. 3*b* and 6*a*). In serine proteases, the well known oxyanion hole is formed from two backbone NH groups, which form hydrogen bonds to the carbonyl group of the substrate and stabilize the negatively charged oxyanion of the intermediate (Hedstrom, 2002). Similarly, in this case the two carbonyl backbones seem to form an amino-binding hole that may enhance the substrate binding, stabilize the intermediate, facilitate ammonium-ion formation and expel the product AROPP.

To demonstrate the importance of the amino-binding hole in BsRibG, a DAROPP analogue, 5-amino-6-ribosylamino-4(3*H*)-pyrimidinone 5′-phosphate (ARPP), with the target amino group substituted by an H atom, was synthesized through inosine triphosphate (ITP) hydrolysis by *E. coli* GCH II. The activity assay was monitored by C_{18} column analysis (Fig. 5*b*). As found for *E. coli* RibD (Magalhães *et al.*, 2008), ARPP could not serve as either a substrate or an inhibitor of BsRibG. This demonstrates a strict requirement for the amino group during binding and thereby suggests an essential role for this amino-binding hole.

4. Discussion

4.1. The essential geometric distortions in the substrate or product

In BsRibG, the carbonyl moiety of the product AROPP is not in the plane of the pyrimidine ring (Fig. 7). Such geometric distortion is also observed in the substrate and product complex structures of mouse CDA (Teh *et al.*, 2006). Another type of distortion is observed in the structures of *E. coli* CDA and *Aspergillus terreus* blasticidin S deaminase (AtBSD) (Xiang *et al.*, 1996, 1997; Kumasaka *et al.*, 2007). The N1—C1′ glycosidic bond is bent out of the pyrimidine plane in order to preserve the coplanarity of the amino/carbonyl group with the pyrimidine. These geometric distortions allow the amino group of the substrate to bind at the amino-binding hole while the carbonyl group of the product coordinates the catalytic zinc ion. Thus, such distortions may also occur in the substrate DAROPP when it binds to BsRibG. Like *E. coli* RibD (Magalhães *et al.*, 2008), GMP, which is a side product in the first step of riboflavin biosynthesis, is unable to serve as either a substrate or an inhibitor of BsRibG (Fig. 5*b*). This may be explained by the distortions mentioned above, which may make it impossible for GMP to bind to BsRibG with significant affinity owing to the restrictions on guanine-ring planarity and the N9—C1′ glycosidic bond. In addition, the six complex structures of AtBSD revealed large positional shifts and changes between tetrahedral and trigonal bipyramidal coordination of the catalytic zinc ion (Kumasaka *et al.*, 2007). The distortions of the substrate and product were proposed to cause structural differences for release of the distortion energy during the substrate–product transition and product release. Therefore, the geometric distortions of the bound substrate and product may play an essential role in the enzyme catalysis.

4.2. Implications relating to RNA-editing deaminases

A comparison of the available crystal structures of CDAs shows that cytosine deaminase, guanine deaminase, dCMP deaminase, bacterial TADA, the C-terminal domain of APOBEC3G and even the large deaminase domain of human ADAR2 share a conserved amino-binding hole (Fig. 8); Ko *et al.*, 2003; Liaw *et al.*, 2004; Almog *et al.*, 2004; Macbeth *et al.*, 2005; Losey *et al.*, 2006; Holden *et al.*, 2008; Furukawa *et al.*, 2009). The consecutive carbonyl backbones are located between the β3 strand and the following short turn prior to the PC*XX*C motif. The β3 strand is located in the centre of the core structure and shares the most highly conserved sequence compared with the other four β-strands (Chen *et al.*, 2006). Most CDA structures could be superimposed well with respect to the β3–αC region (Fig. 8*a*). The orientations of the carbonyl backbones are ensured by the signature proline and the short turn. In contrast, in the tetrameric CDA and BSD an extra ∼5–10 residues have been inserted, which results in one carbonyl backbone moving away; hence, the amino-binding hole is not fully formed (Fig. 8*b*). Recently, structures of the C-terminal deaminase domain of the human antiviral factor APOBEC3G were determined by NMR spectroscopy and X-ray diffraction methods (Holden *et al.*, 2008; Chen *et al.*, 2008; Furukawa *et al.*, 2009). These structures showed large differences in the predicted ssDNA-binding residues. The putative amino-binding hole is still retained in the wild-type

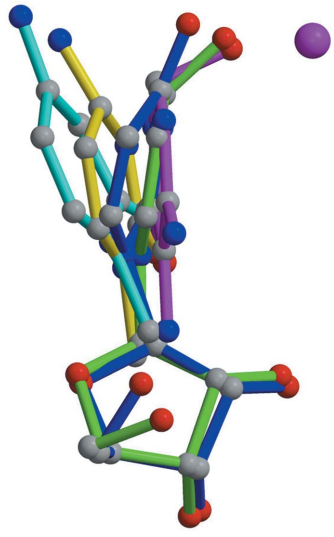


Figure 7
Superposition of substrate and product binding with significant geometric distortions. The products in BsRibG, *E. coli* CDA and mouse CDA are colored magenta, blue and green, respectively, while the substrate or analogue for *E. coli* CDA and mouse CDA are shown in cyan and yellow. The zinc ion is shown as a magenta sphere. Only the pyrimidine moieties of AROPP and cytidine are displayed for clarity.

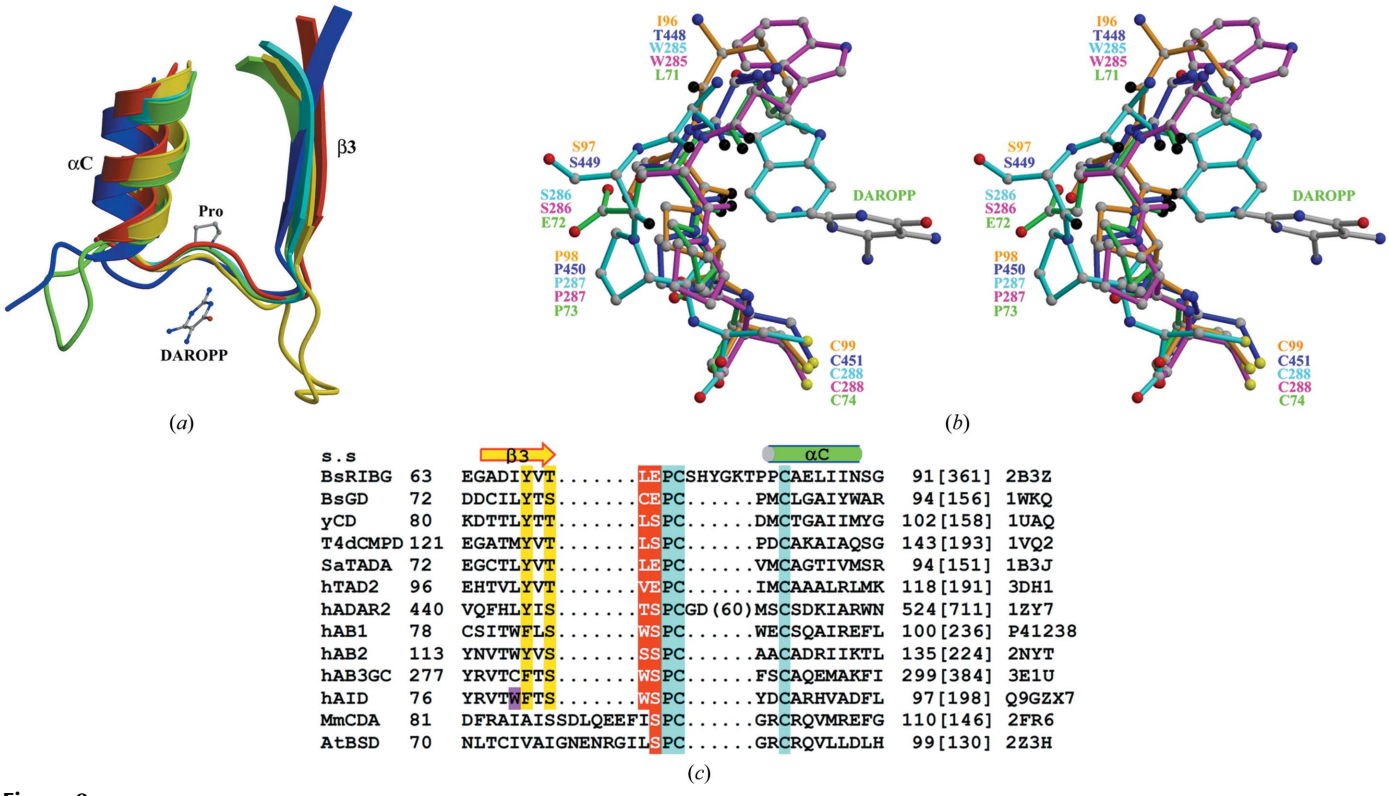


Figure 8
The conserved amino-binding hole. (*a*) Crystal structure superposition of the β3–αC region of BsRibG (green), mouse CDA (yellow), *Staphylococcus aureus* TADA (cyan), the C-terminal domain of human APOBEC3G (hAB3GC; red) and human ADAR2 (blue). The pyrimidine moiety of AROPP and Pro72 in BsRibG are shown as ball-and-stick representations. (*b*) Superposition of the consensus proline and cysteine and the two residues located before the signature in BsRibG (green), mouse CDA (orange), a crystal structure of wild-type hAB3GC (magenta), human ADAR2 (blue) and a solution structure of a mutant hAB3GC (cyan). The two consecutive carbonyl O atoms are highlighted in black. (*c*) Multiple sequence alignment of the β3–αC region. The PDB or GenBank codes are listed in the right column and the protein length is indicated in parentheses. The signature residues are shaded in cyan, while the conserved Tyr/Phe and Ser/Thr in the β3 strand are shaded in yellow. The residues involved in the putative amino-binding hole are shaded in red and Trp80 in human activation-induced deaminase is shaded in magenta.

X-ray and NMR structures, but not in the mutant NMR structure (Fig. 8*b*).

Sequence alignment revealed highly conserved tyrosine (or phenylalanine) and threonine (or serine) residues in the β3 strand (Fig. 8*c*). The structures show that these two residues form close contacts with residues in β1, β4 and αA, and thereby may enhance the stability of the core structure. In the APOBEC members there are two conserved tryptophan residues in this region. The first is located in the middle of the β3 strand (Trp80 in human activation-induced deaminase) and forms extensive hydrophobic contacts with the surrounding residues. As a result, the loss of function of the W80R mutant which is present in hyper IgM syndrome patients may arise from a negative effect on the structural stability of the enzyme (Ta *et al.*, 2003). On the other hand, the second tryptophan at the C-terminus of the β3 strand faces toward the target base (Fig. 8*b*) and may be involved in target-sequence preference (Harris *et al.*, 2002).

4.3. The deamination mechanism in BsRibG and yRib2

Several structural studies of members of the CDA superfamily have demonstrated that they share a conserved zinc-assisted hydrolytic deamination mechanism. Even though BsRibG and yRib2 possess distinct substrate-binding residues (Fig. 6), they share a virtually identical catalytic centre including the zinc ion and the invariant glutamate. They also display similar interaction networks with the pyrimidine ring

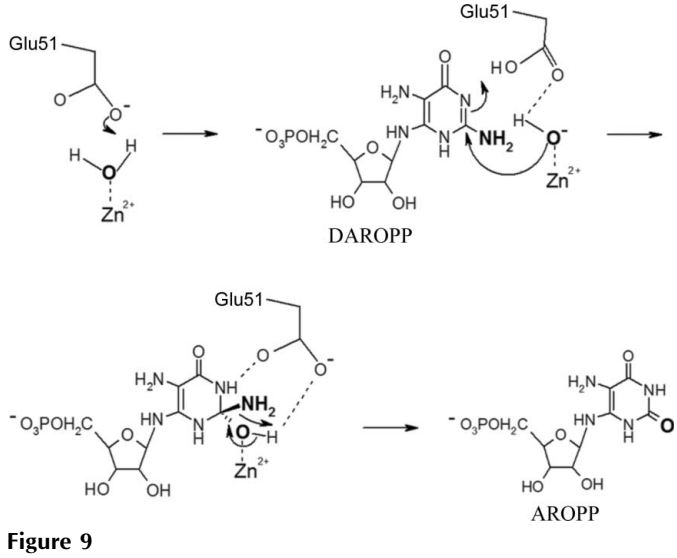


Figure 9
The proposed deamination process for the BsRibG deaminase domain.

bearing the target amino group (Figs. 3*b* and 6*b*). Therefore, it is expected that yRib2 will follow a similar deamination mechanism to that of BsRibG.

On the basis of this BsRibG–AROPP study and previous studies of other CDA members, a deamination mechanism for riboflavin biosynthesis has been proposed (Fig. 9). The catalytic zinc ion and Glu51 interact with the nucleophile water molecule and facilitate its deprotonation. The substrate DAROPP binds to the active-site cavity with its amino group bending out of the pyrimidine ring; this group then binds at the amino-binding hole that involves Leu71 O and Glu72 O (Figs. 3*b*). The R and P loops move towards the substrate to allow direct interactions with the DAROPP ribosyl and phosphate groups. The nucleophile hydroxide ion attacks the pyrimidine C^2, followed by protonation of N^3 to form the intermediate. Glu51 assists in stabilization of the intermediate and proton transfer from O^2H to N^2H_2 then occurs to yield the product. The newly formed carbonyl group of the product AROPP bends out of the pyrimidine plane and ligates the zinc ion. Together with the potentially unfavourable contacts with Glu51, Leu71 O and Glu72 O, such geometric distortion of the carbonyl group may promote product release from the deaminase domain and binding to the reductase domain for sequential reduction.

We thank Dr R. Kirby for critical reading of the manuscript. This study was supported by the National Science Council (NSC100-2311-B010-006-MY3) and a grant from the Ministry of Education, Aim for the Top University Plan. Portions of this research were carried out at the National Synchrotron Radiation Research Center, a national user facility supported by the National Science Council of Taiwan, ROC. The Synchrotron Radiation Protein Crystallography Facility is supported by the National Core Facility Program for Biotechnology.

References

Abbas, C. A. & Sibirny, A. A. (2011). *Microbiol. Mol. Biol. Rev.* **75**, 321–360.

Almog, R., Maley, F., Maley, G. F., Maccoll, R. & Van Roey, P. (2004). *Biochemistry*, **43**, 13715–13723.

Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). *Bioinformatics*, **22**, 195–201.

Bacher, A., Eberhardt, S., Fischer, M., Kis, K. & Richter, G. (2000). *Annu. Rev. Nutr.* **20**, 153–167.

Behm-Ansmant, I., Grosjean, H., Massenet, S., Motorin, Y. & Branlant, C. (2004). *J. Biol. Chem.* **279**, 52998–53006.

Bjelkmar, P., Larsson, P., Cuendet, M., Hess, B. & Lindahl, E. (2010). *J. Chem. Theory Comput.* **6**, 459–466.

Brunger, A. T. (2007). *Nature Protocols*, **2**, 2728–2733.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Chen, S.-C., Chang, Y.-C., Lin, C.-H., Lin, C.-H. & Liaw, S.-H. (2006). *J. Biol. Chem.* **281**, 7605–7613.

Chen, K.-M., Harjes, E., Gross, P. J., Fahmy, A., Lu, Y., Shindo, K., Harris, R. S. & Matsuo, H. (2008). *Nature (London)*, **452**, 116–119.

Chen, S.-C., Lin, Y.-H., Yu, H.-C. & Liaw, S.-H. (2009). *J. Biol. Chem.* **284**, 1725–1731.

Conticello, S. G. (2008). *Genome Biol.* **9**, 229.

DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.

Esnouf, R. M. (1997). *J. Mol. Graph. Mod.* **15**, 132–134.

Fischer, M. & Bacher, A. (2006). *Physiol. Plant.* **126**, 304–318.

Furukawa, A., Nagata, T., Matsugami, A., Habu, Y., Sugiyama, R., Hayashi, F., Kobayashi, N., Yokoyama, S., Takaku, H. & Katahira, M. (2009). *EMBO J.* **28**, 440–451.

Graupner, M., Xu, H. & White, R. H. (2002). *J. Bacteriol.* **184**, 1952–1957.

Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. (2002). *Mol. Cell*, **10**, 1247–1253.

Hedstrom, L. (2002). *Chem. Rev.* **102**, 4501–4524.

Holden, L. G., Prochnow, C., Chang, Y. P., Bransteitter, R., Chelico, L., Sen, U., Stevens, R. C., Goodman, M. F. & Chen, X. S. (2008). *Nature (London)*, **456**, 121–124.

Kaiser, J., Illarionov, B., Rohdich, F., Eisenreich, W., Saller, S., den Brulle, J. V., Cushman, M., Bacher, A. & Fischer, M. (2007). *Anal. Biochem.* **365**, 52–61.

Ko, T.-P., Lin, J.-J., Hu, C.-Y., Hsu, Y.-H., Wang, A. H.-J. & Liaw, S.-H. (2003). *J. Biol. Chem.* **278**, 19111–19117.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Kumasaka, T., Yamamoto, M., Furuichi, M., Nakasako, M., Teh, A.-H., Kimura, M., Yamaguchi, I. & Ueki, T. (2007). *J. Biol. Chem.* **282**, 37103–37111.

Liaw, S.-H., Chang, Y.-J., Lai, C.-T., Chang, H.-C. & Chang, G.-G. (2004). *J. Biol. Chem.* **279**, 35479–35485.

Long, Q., Ji, L., Wang, H. & Xie, J. (2010). *Chem. Biol. Drug Des.* **75**, 339–347.

Losey, H. C., Ruthenburg, A. J. & Verdine, G. L. (2006). *Nature Struct. Mol. Biol.* **13**, 153–159.

Macbeth, M. R., Schubert, H. L., Vandemark, A. P., Lingam, A. T., Hill, C. P. & Bass, B. L. (2005). *Science*, **309**, 1534–1539.

Mack, M. & Grill, S. (2006). *Appl. Microbiol. Biotechnol.* **71**, 265–275.

Magalhães, M. L., Argyrou, A., Cahill, S. M. & Blanchard, J. S. (2008). *Biochemistry*, **47**, 6499–6507.

Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.

Nicholls, A., Sharp, K. A. & Honig, B. (1991). *Proteins*, **11**, 281–296.

Nishikura, K. (2010). *Annu. Rev. Biochem.* **79**, 321–349.

Richter, G., Fischer, M., Krieger, C., Eberhardt, S., Lüttgen, H., Gerstenschläger, I. & Bacher, A. (1997). *J. Bacteriol.* **179**, 2022–2028.

Richter, G., Ritz, H., Katzenmeier, G., Volk, R., Kohnle, A., Lottspeich, F., Allendorf, D. & Bacher, A. (1993). *J. Bacteriol.* **175**, 4045–4051.

Smith, H. C., Bennett, R. P., Kizilyer, A., McDougall, W. M. & Prohaska, K. M. (2012). *Semin. Cell Dev. Biol.* **23**, 258–268.

Ta, V.-T., Nagaoka, H., Catalan, N., Durandy, A., Fischer, A., Imai, K., Nonoyama, S., Tashiro, J., Ikegawa, M., Ito, S., Kinoshita, K., Muramatsu, M. & Honjo, T. (2003). *Nature Immunol.* **4**, 843–848.

Teh, A.-H., Kimura, M., Yamamoto, M., Tanaka, N., Yamaguchi, I. & Kumasaka, T. (2006). *Biochemistry*, **45**, 7825–7833.

Xiang, S., Short, S. A., Wolfenden, R. & Carter, C. W. (1996). *Biochemistry*, **35**, 1335–1341.

Xiang, S., Short, S. A., Wolfenden, R. & Carter, C. W. (1997). *Biochemistry*, **36**, 4768–4774.

Yuvaniyama, J., Chitnumsub, P., Kamchonwongpaisan, S., Vanichtanankul, J., Sirawaraporn, W., Taylor, P., Walkinshaw, M. D. & Yuthavong, Y. (2003). *Nature Struct. Biol.* **10**, 357–365.