Radical-mediated ring contraction in the biosynthesis of 7-deazapurines

Vahe Bandarian\textsuperscript{a,*} and Catherine L Drennan\textsuperscript{b}
\textsuperscript{a}Department of Chemistry, University of Utah, Salt Lake City, UT 84112
\textsuperscript{b}Departments of Biology and Chemistry, Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139

Abstract

Pyrrolopyrimidine containing natural products are widely distributed in Nature. The biosynthesis of the 7-deazapurine moiety that is common to all pyrrolopyrimidines entails multiple steps, one of which is a complex radical-mediated ring contraction reaction catalyzed by CDG synthase. Herein we review the biosynthetic pathways of deazapurines, focusing on the biochemical and structural insights into CDG synthase.

Introduction

Pyrrolopyrimidine functional groups, more commonly referred to as 7-deazapurines, are components of a large class of biological molecules that are found in all domains of life. To date, over 30 deazapurines have been isolated from diverse sources as diffusible metabolites produced by microorganisms in soil and marine environments or hypermodified bases incorporated into tRNA of nearly all organisms (see [1] for a recent review). The widespread distribution of deazapurines suggests that their biosynthetic pathways evolved early and have been maintained because they play significant roles, which remain to be established. Most were isolated on the basis of their herbicidal, antibacterial, antifungal, and antineoplastic activities, suggesting that this scaffold may be particularly suited for incorporation into therapeutic agents.

7-Cyano-7-deazaguanine, commonly referred to as toyocamycin, was the first of the 7-deazapurines described in the literature [2] (Fig 1). The compound was found as a diffusible metabolite in the culture medium of \textit{Streptomyces toyocaensis} and isolated on the basis of its anti-candida activity. The deazapurine-containing hypermodified RNA base, queuosine, was first found in \textit{Escherichia coli} but has since been shown to be present in Asp, Asn, His, and Tyr tRNA of nearly all organisms [3–5]. Early attempts at elucidating the biosynthetic pathways to these compounds utilized radioisotope tracers [6–8] (see Fig. 1 for summary).

\textit{Corresponding author:} University of Utah, Department of Chemistry, 315 S 1400 E, Salt Lake City UT 84112-0850, vahe@chem.utah.edu.

\textit{Publisher’s Disclaimer:} This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Regardless of the metabolite, purines were found to be the starting material, and analysis of labeling patterns showed that the C-2 of the starting purine was retained whereas C-8 was not. The similarity between this distinctive labeling pattern to that observed for folic acid and riboflavin suggested that they may share common biosynthetic steps [9–13]. In addition, the observation that C-1’, C-2’, and C-3’ of the ribose in the proffered purine are incorporated into the deazapurine suggested unprecedented rearrangements [6].

Biosynthesis of Deazapurines

The biosynthetic pathway to the 7-deazapurine core was elucidated by identification of the cluster of genes involved in the biosynthesis of sangivamycin and toyocamycin in \textit{Streptomyces rimosus} [14]. Since the genes for biosynthesis of secondary metabolites in \textit{Streptomyces} tend to be clustered, the search for the biosynthetic pathway focused on identification of a nitrile hydratase activity, which had been shown to convert toyocamycin to sangivamycin. The toyocamycin nitrile hydratase (TNHase) protein was purified and the N-terminal sequences of this heterotrimeric protein were used to identify the gene cluster in a cosmid library of \textit{S. rimosus}. Sequencing a ~13 kbp region surrounding the TNHase genes led to identification of 13 open reading frames which, based on sequence comparisons, appeared to encode the biosynthetic pathway. Independently, four genes of unknown function were shown to be required for the biosynthesis of queuosine in \textit{Acenitobacter calcoaceticus} [15].

Bioinformatic analysis of the \textit{orfs} in the toyocamycin/sangivamycin gene cluster led to two key insights. First, as had been suspected from the radiotracer experiments discussed above, the cluster revealed a link between folic acid and deazapurine biosynthetic pathways. Specifically, a GTP cyclohydrolase I (GCH I) homolog was found in the cluster. GCH I catalyzes the conversion of GTP to 7,8-dihydroneopterin triphosphate (H$_2$NTP) in the first step of folic acid biosynthesis, and entails loss of C-8 and retention of the C-2 of GTP. Second, three of the \textit{orfs} in the cluster are homologous to proteins encoded by genes deemed essential for the biosynthesis of queuosine in \textit{A. calcoaceticus}, suggesting that \textit{Nature} employs similar paradigms for biosynthesis of all pyrrolopyrimidines.

The key steps in the biosynthetic pathway were elucidated by \textit{in vitro} reconstitution of the enzymatic transformations [14,16,17]. The biosynthetic pathway to deazapurines can be divided into two phases. In the first, 7-carboxy-7-deazaguanine (CDG), which is likely the common intermediate to all deazapurines, is formed from GTP (Fig. 2). The second phase in the biosynthetic pathway is different for each deazapurine and involves steps to tailor the CDG to the desired metabolite.

The first phase of the biosynthetic pathway entails conversion of GTP to CDG by three successive enzymatic transformations [16,17]. The first reaction is catalyzed by GCH I, which converts GTP to H$_2$NTP [14,18]. This step is common to the biosynthesis of folic acid, which was expected on the basis of the similarities between the radiotracer experiments [6,7]. Examination of genome databases suggests that the GCH I homolog that is involved in the biosynthesis of folic acid also serves to provide the H$_2$NTP required for the biosynthesis of deazapurines.
In the second step H$_2$NTP undergoes sidechain cleavage catalyzed by 6-carboxytetrahydropterin (CPH$_4$) synthase [17]. CPH$_4$ synthase is homologous to the 6-pyruvoyltetrahydropterin synthase, which catalyzes the second step in the mammalian pathway to biopterin involving conversion of H$_2$NTP to 6-pyruvoyltetrahydropterin (PPH$_4$) [19]. Interestingly, biochemical studies have shown that the CPH$_4$ synthase involved in the biosynthesis of 7-deazapurines also accepts PPH$_4$ and sepiapterin as substrates, but produces CPH$_4$ as the only product [17]. Structural studies suggest that despite significant structural similarities, changing the identities of two amino acids in the active site of CPH$_4$ synthase relative to the mammalian PTPS may explain the differing fates of H$_2$NTP [20].

The third and key ring contraction step to form the pyrrolopyrimidine core is catalyzed by 7-carboxy-7-deazaguanine (CDG) synthase [16]. CDG synthase catalyzes a complex radical-mediated ring contraction reaction to convert CPH$_4$ to CDG. Substantial biochemical and structural evidence is now available to illuminate the mechanism of this fascinating transformation [16,21,22].

The second phase of the biosynthetic pathway entails tailoring CDG to the desired secondary metabolite.

The biosynthetic pathways diverge in the second phase, which involves tailoring CDG. Five steps are involved in tailoring CDG to the hypermodified tRNA base, queuosine (Fig. 3A). In the first, CDG is converted to preQ$_0$ by the action of preQ$_0$ synthetase, which catalyzes the ATP-dependent conversion of the carboxyl moiety of the substrate to a cyano group [16,23]. The ATP is used to activate CDG as an adenylate which, in the presence of ammonia, forms a 7-amido-7-deazaguanine intermediate (ADG) [23]. ADG is subsequently converted to preQ$_0$ in an ATP-dependent manner [23]. In the second step, preQ$_0$ is reduced by an NADPH-dependent reductase to generate preQ$_1$ [24], which is exchanged for guanine in the wobble position of tRNA in the third step [25,26]. This is an unusual exchange reaction that involves formation of a covalent adduct between the ribose and the enzyme, followed by attack with the preQ$_1$ base [27–29]. The fourth step is modification of preQ$_1$-tRNA with a cyclopentanediol epoxide moiety derived from S-adenosyl-L-methionine (SAM) to form epoxyqueuosine (oQ) [30,31]. In the final step the epoxide moiety is reduced by oQ reductase to form queuosine (Q). oQ reductase was identified recently by searching an E. coli knockout library for a strain that was devoid of Q in cellular RNA [32]. Despite their vastly different substrates, oQ reductase is homologous to enzymes that carry out reductive dehalogenation [33,34]. The enzyme has been shown to contain two 4Fe-4S clusters and a cobalamin cofactor, which are all required for activity [35].

With the exception of sangivamycin and toyocamycin (Fig. 3B), the steps involved in tailoring CDG to the secondary metabolites containing 7-deazapurines, have not been determined. The biosynthetic gene cluster for production of sangivamycin and toyocamycin, in addition to a preQ$_0$ synthetase homolog, encodes several nucleotide biosynthesis/salvage enzymes, which are hypothesized to convert preQ$_0$ to toyocamycin [14]. The conversion of toyocamycin to sangivamycin is catalyzed by a cobalt-type nitrile hydratase enzyme [14,36].
The remainder of this review will focus on the current understanding of CDG synthase, which catalyzes the key complex radical-mediated ring contraction reaction in the biosynthetic pathway.

Radical-mediated transformation catalyzed by CDG synthase

CDG synthase is a member of the radical SAM (RS) superfamily [37]. This superfamily was defined on the basis of the sequence motif, CxxxCxxC, which provides thiolato ligands to bind three iron atoms of a 4Fe-4S cluster. The fourth iron coordinates the amino and carboxylate moieties of SAM [38]. The cluster in the +1 oxidation state reductively cleaves SAM to generate 5′-deoxyadenosyl radical (dAdo•), which is subsequently utilized for radical-mediated transformations (Fig. 4A). In most of these enzymes, dAdo abstracts a hydrogen atom to initiate catalysis. It is estimated that >113,000 radical SAM homologs are encoded in bacterial genomes [39].

Interestingly, CDG synthases display important exceptions to the CxxxCxxC consensus motif, whereby homologs with both the canonical and non-canonical (Cx1Cx2C) spacing of Cys residues have been identified [16,22]. A structure of the canonical homolog is not known, but it is expected to bind its FeS cluster as has been observed in all other members of the superfamily [40]. The X-ray crystal structure of the non-canonical homolog from Burkholderia multivorans shows that the 11-amino acid insertion in the CxxxCxxC motif forms a 3_10-helix above the FeS cluster [22]. Also, this non-canonical homolog does not display the typical partial (β/α)6 barrel fold and instead comprises a (β6/α3) barrel.

Activation of RS enzymes requires reduction of the catalytic cluster to the +1 oxidation state. In vitro, low potential reductants such as dithionite are commonly used to activate the protein [41]. However, in vivo it is thought that the electrons are supplied by NADPH via a flavodoxin (Fld)/flavodoxin reductase (Fpr) system (see Fig. 4A). As mentioned above, the reduced Fe-S cluster catalyzes the reductive cleavage of SAM to generate the highly reactive oxidant dAdo•, which initiates the catalytic cycle by abstracting a H-atom from the substrate (Fig. 4B). In a subset of radical SAM enzymes, reductive cleavage of SAM occurs stoichiometrically, whereas in others, SAM plays a catalytic role. CDG synthase utilizes SAM catalytically and at least 10 turnovers can be achieved under optimal conditions with each SAM [21]. However, in CDG synthase and nearly all other radical SAM enzymes, there is an abortive cleavage reaction that occurs when SAM is reductively cleaved and the dAdois quenched by H-atom transfer from a different site. In CDG synthase, the rate of the abortive cleavage reaction with the Bacillus subtilis protein is below the limit of detection when NADPH serves reducing equivalents via FldA/Fpr. In the absence of the complete biological reducing system, the rate is low (~0.3 min⁻¹) and not dependent on the presence or absence of FldA [42].

Although Fld/Fpr are hypothesized to be involved in the activation of RS enzymes and routinely used in vitro to activate RS enzyme, little is known about this interaction at the molecular level. Further confounding this, two flavodoxin homologs are encoded in the genomes of many organisms and only one of the two E. coli Fld homolog, FldA, has been used in RS enzymology to date. A recent kinetic study of the reductive activation of B.
subtilis CDG synthase revealed that one of the two flavodoxin homologs encoded YkuN, is 10-fold more efficient in supporting activation of the enzyme (Fig. 5A) [42]. By comparison, the E. coli Fld homolog was shown to be less efficient in maintaining B. subtilis CDG synthase activity. In these experiments, efficiency is measured as the ratio of maximal activity obtained at saturating Fld ($k_{cat}$) and concentration of Fld that afforded half maximal activation ($K_{Fld}$). Similar results were obtained regardless of whether Fld, which was reduced with dithionite, or with NADPH/Fpr (see Fig. 5B), was used in the assays with B. subtilis CDG synthase. In all cases, Fld was a better mediator than dithionite, suggesting that the differences are likely localized to the Fld•CDG synthase complex. There are many caveats in the interpretation of these results, including the fact that the in vivo concentrations of the various Fld homologs are not known. As the differences in efficiency of activation of CDG synthase by chemical and biological reductants highlights, a better understanding of reductive activation may be necessary if one is to be able to elucidate the reactions catalyzed by this large and growing superfamily of enzymes.

**Catalytic Mechanism of CDG synthase**

The mechanism by which the 6-membered ring of CPH$_4$ undergoes ring contraction has been examined by isotope labeling experiments (Fig. 6) [21]. Upon reductive cleavage, CDG synthase would be expected to carry out H-atom abstraction from C6 or C7 of the substrate. Studies with C-6 and C-7 deuterated isotopologs of CPH$_4$ substrates have demonstrated that the reaction is initiated by H-atom abstraction at C-6. Incubation of CDG synthase with isotopologs of CPH$_4$ where deuterium is located at the 7-proS or 7-proR positions of the substrate, does not lead to labeling of dAdo. In the X-ray crystal structure of CPH$_4$ bound to CDG synthase, the 5’-position of the cofactor is within 3.4 Å of the C-6 of the substrate, poised for H-atom abstraction upon reductive cleavage of the cofactor [22] (see Fig. 6 insert).

Abstraction of a H-atom from C-6 of CPH$_4$ leads to an initial radical intermediate, which would be stabilized by delocalization. Although there is no structural or spectroscopic evidence for this intermediate, the structure of the 6-carboxypterin, which is a substrate analog, provides an indication of localization as it mimics the planar arrangement at C-6 [22] (see Fig. 6 insert). The mechanism by which this intermediate undergoes ring contraction is not known, but there are two reasonable possibilities. One may envision that the reaction would proceed via an aziridine-like intermediate, where the unpaired spin density would be stabilized by delocalization. Alternatively, the rearrangement may proceed through a ring opening followed by closure by 5-exo-trig to generate the 5-membered ring. Computational studies favor the aziridine-like intermediate in the rearrangement [43]. In both cases, the resulting radical would be quenched by H-atom abstraction of an H-atom from the dAdo to form the gem-aminocarboxylate intermediate and allow regeneration of the cofactor. Although a structure of this complex is not available, its position can be modeled on the basis of the site of binding of CDG (see Fig. 6 insert).

The conversion of the gem-aminocarboxylate intermediate to CDG requires elimination of ammonia. Studies with the isotopologs of CPH$_4$ have shown that when deuterated at C-7, the label at the 7-proR position is selectively retained in the CDG product [21]. This
observation suggests that the elimination of ammonia and aromatization to form CDG occur on the enzyme. Computational studies suggest that elimination of the ammonia to form the CDG product is the rate-determining step [43]. It is possible, for example, that the exocyclic amino group of the pyrimidine ring could be used to eliminate ammonia. Such an intermediate would have an acidic proton at C-7 that can be abstracted by an appropriate base to generate the product. Indeed, in the structure of the product complex, a Glu side chain is within distance from the appropriate face of the substrate for abstraction (see Fig. 6 insert).

Superposition of the structures of the substrate, intermediate analog, and product complexes of CDG highlight the fact that there is minimal movement in the active site of the protein as the substrate undergoes activation and conversion to product (see Fig. 7 for overlay). This design principle is an emerging theme in RS enzymes because the highly reactive intermediates can do side-reactions if they are not controlled [44,45]. One may quip that while technically these enzymes produce “free radical intermediates”, there is nothing “free” about the intermediates!

A curious observation that was made in the course of the biochemical studies of CDG synthase was that in addition to an FeS cluster, CDG synthase also requires a Mg$^{2+}$ for catalytic activity [21]. In addition to revealing details of binding of the substrate, the structures of the B. multivorans homolog also provide insights into the role of the required magnesium divalent cation [22]. Close examination of the binding interactions between CPH$_4$ and the enzyme shows that while the substrate makes several contacts with the enzyme, the magnesium divalent cation “templates” the substrate-binding site (Fig. 8). The Mg$^{2+}$ interacts with the carboxylate oxygen as well as the carbonyl oxygen in the pyrimidine ring. By contrast, the only contact between the divalent cation and the enzyme is an interaction with the hydroxyl sidechain of a threonine residue. It is worth noting that the substrate makes nearly as many interactions with the bound divalent cation as with protein side chains, suggesting that the divalent cation serves as a key binding determinant. A role for the Mg$^{2+}$ beyond positioning of the substrate in the active site is not known. It has been suggested that it may serve as a Lewis acid to facilitate the deamination of the substrate.

**Future directions**

The radical-mediated ring contraction catalyzed by CDG synthase is complex and affords the ideal platform to explore the structural and biochemical details of reductive activation and catalysis. CDG synthase is only one of several novel enzymes that were discovered in the course of studies of biosynthesis of 7-deazapurines, and future studies on the biosynthesis of the remaining pyrrolopyrimidines may lead to discovery of additional novel transformations.

**Acknowledgements**

Research reported in this publication was supported by the National Institute of General Medical Science of the National Institutes of Health under award number R01 GM72623. Additional support from a Career Award in Biomedical Sciences from the Burroughs Wellcome Fund (to V.B.) is also acknowledged. C.L.D. is a Howard Hughes Medical Investigator.
References and recommended reading

20. Miles ZD, Roberts SA, McCarty RM, Bandarian V. Biochemical and structural studies of 6-carboxy-5,6,7,8-tetrahydropterin synthase reveal the molecular basis of catalytic promiscuity within the tunnel-fold superfamily. J. Biol. Chem. 2014; 289:23641–52. [PubMed: 24990950]


37. Sofia HJ, Chen G, Hetzler BG, Reyes-Spindola JF, Miller NE. Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms:


43. Zhou W, Liu Y. Ring contraction catalyzed by the metal-dependent radical SAM enzyme: 7-carboxy-7-deazaguanine synthase from B. multivorans. Theoretical insights into the reaction mechanism and the influence of metal ions. ACS Catalysis. 2015; 5:3953–3965. *This is a comparative study of reductive activation of B. subtilis CDG synthase by biological and non-biological reductants.


Highlights

- Pyrrolopyrimidines are widely distributed in Nature.
- Biosynthesis of 7-deazapurine core of all pyrrolopyrimidines entails a radical-mediated ring contraction
- CDG synthase catalyzes a complex radical-mediated ring contraction
- CDG synthase is a member of the radical SAM superfamily
Figure 1.
Labeling patterns observed in the biosynthesis of folic acid and deazapurines. In both cases, C-2 (red sphere) of the starting purine is retained but C-8 (grey sphere) is lost. In addition, C-1’, C-2’, and C-3’ (blue spheres) of the proffered purine ribose become incorporated into the final product.
Figure 2.
Core steps in the biosynthesis of 7-deazapurines are catalyzed by the successive actions of three enzymes.
Figure 3.
Tailoring steps from CDG to (A) the hypermodified tRNA base queuosine and (B) the secondary metabolites toyocamycin and sangivamycin.
Figure 4.
Activation of RS enzymes requires reduction of the [4Fe-4S] cluster to the +1 oxidation state. In vivo, NADPH is thought to supply the necessary reducing equivalents via Fpr/Fld (A). Once reduced, the RS cluster catalyzes the reductive cleavage of SAM to generate a dAdo•, which initiates catalysis by H-atom abstraction from the substrate (B).
Figure 5.
Reductive activation of B. subtilis CDG synthase with Fld homologs from E. coli and B. subtilis. Both the biological reducing system NADPH/Fld/Fpr (A) and dithionite/Fld (B) are able to activate CDG synthase. Of the Fld homologs, YkuN is able to maintain activity of CDG synthase at significantly lower concentrations than YkuP or Fld.
Mechanism of CDG synthase. Radical-mediated ring contraction is initiated by H-atom abstraction at C-6 (hydrogen in light blue) of the substrate, and product is formed by stereoselective proton abstraction from a putative gem-aminocarboxylate intermediate. Lower inserts show active sites from the structures of CPH$_4$ (PDB:4NJI), 6-CP (PDB: 4NJG), and CDG (PDB: 4NJK) bound near the [4Fe-4S] of CDG synthase. The structure below the gem-aminocarboxylate intermediate is a model based on how CDG binds to CDG synthase. Glu 116 is also shown in this panel. Colors for the structures: Fe in rust, S in yellow, C in green, N in blue, O in red, and modeled H in white. The 7-proS and 7-proR hydrogens are shown in black and red spheres, respectively.
Figure 7.
CPH₄, CDG and 6-CP are all bound in the same manner with respect to the SAM-bound [4Fe-4S] cluster in the active site of CDG synthase. Color of the FeS cluster and SAM are as described as in Fig. 6.
Figure 8.
Magnesium divalent cation (orange sphere) serves as a major binding determinant for CPH$_4$ in the active site of CDG synthase. The substrate makes three contacts to the Mg$^{2+}$. A Thr sidechain from the protein is also a ligand. The C-terminal carboxylate and Arg27 are also involved in binding the substrate. Colors are as described in Fig. 6 with the hydrogen that is abstracted (white) modeled into the structure.