

Published in final edited form as:

J Biol Chem. 2005 June 17; 280(24): 23000–23008. doi:10.1074/jbc.M501534200.

Structure and function of both domains of ArnA, a dual function decarboxylase and a formyltransferase, involved in 4-amino-4-deoxy-L-arabinose biosynthesis

Gareth J. Williams¹, Steven D. Brezeale², Christian R. H. Raetz², and James H. Naismith^{1,*}

¹Centre for Biomolecular Science, The University, St. Andrews KY16 9ST, UK

²Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710

Abstract

Modification of the lipid A moiety of lipopolysaccharide by the addition of the sugar, 4-amino-4-deoxy-L-arabinose (L-Ara4N), is a strategy adopted by pathogenic Gram-negative bacteria to evade cationic antimicrobial peptides produced by the innate immune system. L-Ara4N biosynthesis is therefore a potential anti-infective target, as inhibiting its synthesis would render certain pathogens more sensitive to the immune system. The bifunctional enzyme ArnA, which is required for L-Ara4N biosynthesis, catalyzes the NAD⁺-dependent oxidative-decarboxylation of UDP-glucuronic acid to generate a UDP-4'-keto-pentose sugar, and also catalyzes transfer of a formyl group from *N*-10-formyltetrahydrofolate to the 4'-amine of UDP-L-Ara4N. We now report the crystal structure of the N-terminal formyltransferase domain in a complex with uridine monophosphate and *N*-5-formyltetrahydrofolate. Using this structure we identify the active site of formyltransfer in ArnA including the key catalytic residues N102, H104, and D140. Additionally, we have shown that residues S433 and E434 of the decarboxylase domain are required for the oxidative-decarboxylation of UDP-GlcUA. An E434Q mutant is inactive suggesting that chemical rather than steric properties of this residue are crucial in the decarboxylation reaction. Our data suggests that the decarboxylase domain catalyzes both hydride abstraction (oxidation) from the C-4' position and the subsequent decarboxylation.

Keywords

site directed mutagenesis; X-ray crystallography; drug design; LPS biosynthesis; polymyxin resistance

Introduction

The outer leaflet of the outer membrane of Gram-negative bacteria predominantly consists of the lipopolysaccharide (LPS) molecule. This immunogenic glycolipid is anchored to the membrane by lipid A, which is the active moiety causing many of the pathophysiological effects associated with Gram-negative sepsis (1). The innate immune response activated by lipid A includes production of cationic antimicrobial peptides (CAMPs), cytokines, clotting factors and other immunostimulatory molecules (2-5). Severe sepsis can lead to a shock syndrome, which is brought on by high levels of cytokines and procoagulants and in some cases causes organ failure and even death (6). CAMPs are small, positively charged peptides, and their antimicrobial properties arise through their interactions with the lipid A moiety (7-9). These peptides interact electrostatically with the negative groups of the LPS,

*To whom correspondence should be addressed..

then traverse the outer membrane, forming pores that result in permeabilisation and cell death (10,11). Many Gram-negative bacteria have evolved resistance to CAMPs by modifying the negative groups of LPS to reduce the net negative charge, thus reducing the attraction to CAMPs (12,13). These modifications include acetylation of the O antigen and addition of phosphoethanolamine and/or 4-amino-4-deoxy-L-arabinose (L-Ara4N) to the 1' and 4' phosphates of lipid A (13-17).

Modification of lipid A by L-Ara4N in *Salmonella typhimurium* can be abolished through mutations in the *pmrE/ugd* gene or the *pmrF* operon, rendering these cells sensitive to polymyxin B (a CAMP type antibiotic) (12,18). The *pmrF* operon encodes 7 genes, which are co-transcribed under PmrA control, that are required for polymyxin resistance, with the exception of *pmrM* (12,18). The *pmrF* operon and *pmrE* gene are also conserved in many other pathogenic bacteria that are known to modify their lipid A with L-Ara4N including; *Escherichia coli* K-12, *Yersinia pestis* and the major cystic fibrosis pathogen *Pseudomonas aeruginosa* (12,13,18,19). The sequence similarity of the enzymes encoded by the *pmrE* and *Floci* to enzymes of known function led to a proposed pathway for the biosynthesis of the L-Ara4N moiety and its attachment to lipid A (Figure 1) (14,20), which has since been largely validated by Raetz and coworkers (21-23). The first step is the conversion of UDP-glucose (UDP-Glc) to UDP-glucuronic acid (UDP-GlcUA), catalyzed by the well characterized enzyme UDP-Glc dehydrogenase. UDP-GlcUA is then oxidatively decarboxylated by ArnA, in the first unique step of the pathway, giving a UDP-4'-keto-pentose (21). This 4'-keto sugar is subsequently converted to a 4'-amino sugar, UDP-L-Ara4N, by the pyridoxal phosphate-dependent enzyme ArnB (22,24). ArnA then catalyzes a second reaction, the transfer of a formyl group from *N*-10-formyltetrahydrofolate (*N*-10-fTHF) to the 4'-amine of UDP-L-Ara4N yielding UDP-L-Ara4FN. The membrane bound protein ArnC catalyzes the transfer of the L-Ara4FN moiety from UDP to undecaprenyl-phosphate, generating Und-P-L-Ara4FN (25). The ArnC enzymatic product is then deformylated by ArnD and transported to the outer surface of the inner membrane by a process that is not yet fully understood. The deformylated L-Ara4N moiety is then transferred from undecaprenyl-phosphate to lipid A by ArnT (23) (Figure 1).

Sequence analysis of ArnA reveals two discrete catalytic domains. The N-terminal formyltransferase domain is homologous to *N*-10-formyltetrahydrofolate; L-methionyl-tRNA *N*-formyltransferase (FMT) and to glycnamide ribonucleotide transformylase (GART). The mechanism of transfer of the formyl group to a free amine group, catalyzed by the FMT and GART enzymes, is quite well understood (26). However, details of the active site structure of the ArnA formyl transferase domain are unknown. The C-terminal domain of ArnA is homologous to UDP-galactose-4-epimerase (GalE), dTDP-glucose-4,6-dehydratase (RmlB) and other short chain dehydrogenase (SDR) enzymes that oxidize the C4' position of sugar nucleotides. The structure of the decarboxylase domain has recently been reported, and it confirms that the enzyme belongs to the SDR class of enzymes, although no biochemical characterization was reported (27). A key question about the ArnA C-terminal decarboxylase domain is the mechanism of decarboxylation. The formation of the keto sugar is well understood from studies of RmlB enzyme. However, whether the decarboxylation is spontaneous or catalyzed by ArnA is not known.

The enzymes involved in the biosynthesis of L-Ara4N are all potential therapeutic targets for the production of new anti-infectives. ArnA is a good target because of its involvement in two steps of this pathway. The well known inhibitory effects of deaza analogues of folic acid on GART activity make the formyltransferase domain particularly interesting (26,28). Therefore an understanding of the mechanism of action and the structure of these two domains will be useful for inhibitor design. Here we report the structure of the N-terminal formyltransferase domain as a complex with UMP and a substrate analog *N*-5-

formyltetrahydrofolate (*N*-5-fTHF), allowing identification of the active site. We have also determined the structure of the C-terminal decarboxylase domain and present biochemical evidence for the involvement of S433 and E434 in the decarboxylation reaction.

Experimental procedures

Overexpression of native and selenomethionine decarboxylase and formyltransferase domains of ArnA

Full length protein was cloned and expressed but failed to yield satisfactory crystals. As a result we chose to work on the separate domains. The N-terminal domain of ArnA from *E. coli* (residues M1 to N305) and the C-terminal domain (T316M to S660) were cloned into pET24b to produce the pET-FT and pET-DH plasmids respectively. *E. coli* BL21DE3 cells were transformed with either the pET-DH or pET-FT plasmids for overexpression of the decarboxylase and formyltransferase domains respectively. Single colonies were grown in 10 ml LB supplemented with 50 µg/ml kanamycin overnight. Overnights were used to inoculate 500 ml of the same LB or 250 ml M9 media, supplemented with 50 µg/ml kanamycin, in 2 l flasks for growth of native or selenomethionine modified (Se-Met) protein respectively. For overexpression of native protein, cells were grown to an OD₆₀₀ of 0.6-0.8 at 37 °C and then induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside at 18 °C overnight. The purification protocols of the two native proteins were very similar; supernatant containing target protein was applied to an HQ anion exchange affinity column (Applied Biosystems) and bound proteins were eluted with an increasing concentration of NaCl (0.05-1 M). NaCl was added to a final concentration of 4.5 M to fractions containing the target protein and then applied to an HP hydrophobic column (Applied Biosystems). The target protein was eluted with a decreasing NaCl gradient, dialyzed into 50 mM HEPES pH 8.0 containing 250 mM NaCl, and further purified by S200 gel filtration chromatography (Amersham Biosciences). Each step of purification was monitored by SDS-polyacrylamide gel electrophoresis. After the gel filtration step both proteins were judged, by coomassie stained gels, to be pure. The decarboxylase and formyltransferase proteins were concentrated to 5 mgml⁻¹ and 8 mgml⁻¹ respectively, and dialyzed into 25 mM HEPES pH 7.5, 50 mM NaCl and 2 mM DTT for crystallization trials. Standard protocols were employed for Se-Met expression (29). Purification of the Se-Met protein is identical to that of the native protein but all buffers also contained 7 mM DTT. Mutants of the decarboxylase domain were purified identically to the native protein.

Association of domains with each other and with ArnB Association of domains with each other and with ArnB

To investigate whether the isolated domains interact with each other, the purified domains, at approximately 3 mgml⁻¹ (approx 0.075 mM), were mixed and incubated for 1 hour at room temperature in buffer containing; 50 mM HEPES pH 7.5, 250mM NaCl. The mixture was applied to an S200 gel filtration column and the elution of proteins was followed by A₂₈₀. No material eluted at the same point as the full length protein which was used as a standard. We repeated the experiment adding 5 mM UDP-GlcUA, 5 mM NAD⁺ and *N*-5-fTHF (Sigma) to the buffer but again no peak corresponding to a complex was observed. We also performed the same analysis with purified ArnB (2 mgml⁻¹) incubating individually with both isolated domains and full length ArnA in 50 mM HEPES pH 7.5, 250 mM NaCl. We repeated the full length ArnA ArnB experiment with 5 mM glutamic acid (Sigma) and 5 mM pyridoxal 5'-phosphate (Sigma) added to buffer. We did not observe any complex formation.

To detect formation of a complex between His-tagged ArnB and ArnA, a pull-down experiment using the Penta•His antibody (Qiagen) was done as described by the

manufacturer. Briefly, purified ArnB-N6H (450 ng) (22) was incubated with a cell-free extract overexpressing ArnA (750 ng) (21) at 4 °C for 90 minutes with shaking in 400 µl of buffer containing 50 mM HEPES pH 7.5 and 10 % glycerol. Nontagged proteins were precipitated with 40 µl of Protein-G agarose (Sigma) before the 2 µg of Penta•His antibody was added. Fresh Protein-G agarose was then added to precipitate ArnB-N6H. The precipitated Protein-G agarose was washed twice with 100 µl of 50 mM HEPES, pH 7.5. Proteins samples from each step were resolved by 10% SDS-PAGE and analyzed by silver staining.

Crystallization and data collection of native and Se-Met formyltransferase protein

Protein was incubated overnight at 4 °C in the presence of 5 mM *N*-5-fTHF and 5 mM UDP (Sigma) before crystal screens were set up. Crystallization of the native protein was tested against commercially available screens (Hampton; screens 1 and 2, index, peg/ion and Wizard 1 and 2), using the sitting drop vapor diffusion method at 20 °C. Drops contained 1 µl protein with 1 µl precipitant, against a well of 100 µl precipitant. Good quality crystals were obtained in 0.1 M ammonium acetate, 0.1 M Bis-tris pH 5.5, 17 % PEG 10000, using a drop containing 4.5 µl protein and 4.5 µl precipitant against a reservoir of 100 µl precipitant. Se-Met protein was screened against sparse matrix screens from Hampton at 20 °C and 4 °C. Good quality crystals formed at 4 °C in 0.22 M ammonium fluoride and 22 % PEG 3350, with drops containing 3 µl protein and 3 µl precipitant against a well of 100 µl mother liquor. Crystals were cryocooled to 100 K using a cryoprotectant of mother liquor containing 20 % glycerol. Data to 1.2 Å was collected on native crystals at ESRF beamline ID14-2; 240 0.5 ° oscillation images were collected, using an exposure time of 3 s and a crystal to detector distance of 91 mm. A Se-Met crystal was used to collect MAD data to 1.6 Å on a MAR-CCD detector, at three wavelengths, on the ESRF beamline ID14-4. Data were indexed and integrated in MOSFLM (30) and scaled in SCALA (31). Merging of the data and analysis of the systematic absences identified the space group as $P2_12_12_1$, with cell dimensions $a=67.2$ Å, $b=90.0$ Å, $c=97.9$ Å, $\alpha=\beta=\gamma=90$ °.

Crystallization and data collection of native and Se-Met decarboxylase protein

Good quality crystals were obtained with a precipitant solution of 3.2 M NaCl, 0.1 M Bis-tris pH 5.2, using a drop containing 4 µl protein and 4 µl precipitant equilibrated against a reservoir of 100 µl precipitant. Se-Met protein crystallized in similar conditions. The crystals were cryocooled to 100 K using a cryoprotectant of 4 M sodium formate. Data to 2.3 Å were collected on native crystals on beamline ID14-2 at the ESRF using an exposure time of 2 s and crystal to detector distance of 217 mm. Se-Met crystals were used to collect data to 3.35 Å on a MAR-CCD detector, at three wavelengths around the Se absorption edge, on beamline BM14 at the ESRF. Data were treated similarly to the formyltransferase domain data. Merging of the data and analysis of systematic absences determined the spacegroup as $P4_132$, with cell dimensions $a=b=c=149.4$ Å, $\alpha=\beta=\gamma=90$ °.

Structure determination and refinement of the decarboxylase and formyltransferase proteins

SOLVE (32) located 4 selenomethionines for the decarboxylase data and 8 selenomethionines for the formyltransferase data. Phases were improved with RESOLVE (33) and extended in DM (34), as implemented in the CCP4 program suite (35), to 2.3 Å and 1.2 Å, using the native decarboxylase and formyltransferase data respectively. Automatic electron density interpretation by ARP/WARP (36) was carried out, and this traced the majority of each structure. Models were completed by manual intervention using O (37) and COOT (38). Refinement of each structure was carried out using REFMAC5 (39), initially with rigid body refinement and later with restrained refinement. Electron density showed the presence of UMP and *N*-5-fTHF in the formyltransferase structure. These compounds were

fitted to the density, and libraries were made using PRODRG (40) for inclusion in refinement. The decarboxylase and formyltransferase structures were refined to $R=18.6\%$, $R_{\text{FREE}}=23.0\%$ and $R=13.4\%$, $R_{\text{FREE}}=15.7\%$ respectively. Crystallographic statistics are shown in Table 1.

Mutagenesis and assay of the decarboxylase domain

Site-specific mutations S433A, E434A and E434Q were introduced into the decarboxylase domain. These mutants were made according to the QuikChange® protocol, using the ArnA-DH plasmid, *pfu* DNA polymerase and the following oligonucleotides:

S433A	3'-CGTCAACTGCAGAAGTTTATGGG-5'
	3'-CCATAAACTTCTGCAGTTGACG-5'
E434A	3'-CGTCAACTTCAGCAGTTTATGGG-5'
	3'-CCATAAACTGCTGAAGTTGACG-5'
E434Q	3'-CCCGTCAACTTCACAGTTTATGGG-5'
	3'-CCATAAACCTGTGAAGTTGACGGG-5'

Mutations were confirmed by DNA sequencing and the proteins overexpressed and purified in the same way as wild type protein. Circular dichroism and fluorescence spectra analysis were identical for wild type and mutant proteins. Wild type and mutant ArnA decarboxylase enzyme activities were assayed by monitoring the rate of NADH production at A_{340} . The assays contained 50 mM HEPES pH 7.5, 300 mM NaCl, 2 mM DTT, 0.2 mgml⁻¹ BSA, 8 mM NAD⁺, 4.5 mM UDP-GlcUA and either 0.5 μ M wild type or 2.5 μ M mutant protein. Protein concentration was judged by BCA assays and all reactions were carried out at 30 °C. K_m values were determined by varying the amount of the initial UDP-GlcUA or NAD⁺ from $1/5 \times K_m$ to $5 \times K_m$. The results for these assays are shown in Table 2. The mutants were further probed by incubation with substrate and analysis by high pressure liquid chromatography (HPLC).

Results

Association of domains with each other and with ArnB

The full-length ArnA protein contains two discrete catalytic domains that are both required for the L-Ara4N modification of lipid A. As shown in Figure 1 the C-terminal decarboxylase domain of ArnA catalyzes the oxidative decarboxylation of UDP-GlcUA to yield UDP-Ara4O, which is then converted to UDP-L-Ara4N by a separate enzyme, ArnB. To synthesize UDP-L-Ara4FN, the N-terminal formyltransferase domain of ArnA then catalyzes the transfer of a formyl group from *N*-10-fTHF to the 4'-amine of UDP-L-Ara4N. Although fusion proteins are common in biology, it is unusual for a bifunctional enzyme to require an intervening enzymatic step. We therefore investigated if the purified discrete domains of ArnA form a complex by themselves, or together with ArnB. To this end, the ArnA domains were combined and passed over an S200 gel filtration column in the presence and absence of cofactors. Only the individual domains were observed, heterodimer formation was not seen. We conservatively estimate that the lower level of detection was about 0.06 mg/ml (1 μ M), correlating to a lower limit of the K_d as 5 mM. A similar analysis was performed by incubating purified ArnB with the N-terminal formyltransferase domain and the C-terminal decarboxylase domain of ArnA or the full length ArnA protein. Again, formation of a complex between ArnB and ArnA (discrete domains or full length protein) was not observed, suggesting that under these conditions the K_d for a complex would be in the millimolar range if not considerably higher. Furthermore, we were not able to detect

formation of a complex by co-precipitation experiments with His-tagged ArnB and ArnA, using the Penta•His antibody in combination with Protein G-agarose.

Overall structure of the formyltransferase domain

The crystal structure of the formyltransferase domain of ArnA has been solved to 1.2 Å in complex with *N*-5-fTHF and UMP. Crystals were grown in the presence of UDP, but only UMP was observed in the active site. The formyltransferase domain of ArnA crystallizes as a dimer and the overall structure, shown in Figure 2a, is similar to that of FMT. The ArnA formyltransferase domain structure consists of two subdomains; an N-terminal subdomain (residues M1-K181) and a C-terminal subdomain (residues F207-N305), which are linked by a loop (residues H182-S206). The N-terminal subdomain is folded into a 7-stranded β -sheet with 5 parallel sheets (β 1-5) and 2 anti-parallel sheets (β 6-7), flanked by a total of 6 α -helices, forming a version of the classic Rossmann fold. This subdomain is similar in structure to GART (26) and the N-terminal domain of FMT (41). The C-terminal subdomain of the formyltransferase domain is similar in structure to the C-terminal domain of FMT and is folded into 2 large β -sheets (β 8 and β 9), 2 small β -sheets (β 10 and β 11) with 2 α -helices (α 7 and α 8). The β -sheets fold to form a small β -barrel, which is flanked on either side by one of the helices, with α 7 located parallel to α 5 of the N-terminal subdomain (Figure 2a).

The active site of the ArnA-formyltransferase domain is located in the same position as in GART and FMT, consistent with all three enzymes utilizing *N*-10-fTHF as the formyl donor. In the structure presented here (Figure 2a) *N*-5-fTHF binds adjacent to β 4 and β 5, and is held in place by hydrogen bonds between the bicyclic ring and two loops, one between β 4 and α 4 (residues I88 and H86) and the other between β 6 and β 7 (residues V136 and D140). Hydrophobic interactions between *N*-5-fTHF and residues L87 (on loop between β 4 and α 4), N102 (on β 5), M135 and A139 (on loop between β 6 and β 7), also assist in the binding of the cofactor analogue. Figure 3a shows these interactions in more detail. Analysis of the thermal factors of the atoms of *N*-5-fTHF indicates the bicyclic ring is very well ordered, while the benzoyl ring is less ordered and the glutamate displays the least order. This is in agreement with the proposal of Almasy *et al.* that the bicyclic ring is more tightly bound than the benzoyl or glutamate moieties in GART (26,41).

UMP is bound mainly by the N-terminal domain and is adjacent to *N*-5-fTHF (Figure 2a and 3a). Y42 is hydrogen bonded to O2P and R85 to O1P of the UMP phosphate, with a further hydrogen bond between O4 of uracil and R201 (Figure 3a). Hydrophobic interactions occur between uracil and N118, T202 and W228. R201 and T202 are part of the loop connecting the N- and C-terminal subdomains and W228 is part of α 7 of the C-terminal subdomain (Figure 3a). These are the only interactions that occur outside of the N-terminal subdomain between the formyltransferase domain and its cofactor or substrate.

Figure 4 shows the sequence alignment between the formyltransferase domain of ArnA, GART and FMT. Highlighted are N102, H104 and D140 (in ArnA), the three conserved residues thought to be involved in catalysis. Their conservation suggests a similar mechanism of formyl transfer as described by Almasy *et al.* (26), with the major differences between the proteins being how they bind their respective substrates UDP-L-Ara4N (ArnA), β -GAR (GART) and M-tRNA (FMT). Although we have been unable to obtain a true substrate co-crystal, the location of UMP and the spatial conservation of the key catalytic residues allow identification of the ArnA formyltransferase active site (Figure 3a).

The decarboxylase domain of ArnA

We solved the apo structure of the decarboxylase domain of ArnA to a resolution of 2.3 Å (Figure 2b) and based our biochemical studies on this structure prior to the publication by

Gatzea-Topalova *et al* (27) describing the structure of the decarboxylase domain to 2.4 Å. The structures are essentially identical. The decarboxylase domain is similar in overall structure to RmlB (42) and GalE (43). The structure can be split into two functional domains; a C-terminal sugar nucleotide binding domain and an N-terminal NAD⁺ binding domain. A cleft is formed between the two functional domains, allowing NAD⁺ and UDP-GlcUA access to the active site.

Co-crystallization and soaking experiments failed to produce a complex of the ArnA decarboxylase domain with NAD⁺ and/or UDP-GlcUA or substrate analogues (UDP-Glc and UDP-galactose). ArnA is a member of the SDR superfamily which is characterized by a catalytic triad of T/S (T in ArnA and RmlB), Y and K. These three residues and NAD(P)⁺ are required for hydride transfer between the substrate and the nicotinamide cofactor, the defining chemical reaction catalyzed by all SDR enzymes. Reduction of NAD⁺ is consistent with the first step in the ArnA decarboxylase mechanism, the oxidation of UDP-GlcUA to 4'-keto-UDP-GlcUA. This oxidation is very similar to the first step of the RmlB mechanism, the oxidation of the dTDP-Glc to 4'-keto-dTDP-Glc. The spatial requirements of hydride transfer from the substrate to NAD⁺ and the positioning of the catalytic triad have been established for RmlB and are quite stringent (42,44). We have superimposed the RmlB structure containing NAD⁺ and dTDP-Glc onto ArnA, and used this to generate a conceptual model of UDP-GlcUA at the active site of ArnA (Figure 3b). In addition to hydride transfer, the SDR superfamily catalyzes a number of other chemical transformations; these are usually dependent upon the presence of oxidized (activated) product. In RmlB, E135 abstracts a proton from C5' and D134 protonates 6'OH, resulting in the elimination of water (42). Our superposition of RmlB and ArnA shows that the ArnA residues S433 and E434, which are conserved throughout all ArnA decarboxylase domains, superimpose with these two key RmlB residues D134 and E135. Gatzea-Topalova *et al* hypothesized, but did not test, a role for the conserved S433 in the reaction as well a role for substrate recognition by R619 (27). They suggested that E344 served to block rotation of the oxidized 4'-keto intermediate. To probe the mechanism further, we mutated S433 and E434 to determine whether this would affect the decarboxylation reaction. As neither residue is conserved across the SDR superfamily, we did not expect these mutations to affect the oxidation step in the reaction (Figure 1).

Biochemical studies

The ArnA decarboxylase domain and the S433A, E434A and E434Q mutant decarboxylase enzymes were assayed for their ability to reduce NAD⁺ to NADH in the presence of UDP-GlcUA, the results of which are shown in table 2. The E434Q mutant has no activity and the specific activities of the S433A and E434A mutants are approximately 40- and 100-times lower than that of the native enzyme, respectively. These mutations do not significantly perturb the K_m for the substrate. The assay only measures the first step of the reaction, oxidation of UDP-GlcUA to 4'-keto-UDP-GlcUA, which we did not expect to be affected by these mutations. Furthermore, in wild-type, S433A, and E434A reactions only the decarboxylated product, UDP-Ara4O, and the substrate UDP-GlcUA are observed by HPLC. The putative intermediate, 4'-keto-UDP-GlcUA, was not seen in any reaction. Instead it is stoichiometrically decarboxylated by the mutant and native enzymes to form the 4'-keto-pentose, UDP-Ara4O.

Discussion

The ArnA protein has two recognizable domains, an N-terminal formyltransferase and a C-terminal decarboxylase, which are both required for the L-Ara4N modification of lipid A and polymyxin resistance. These domains were expressed, purified, crystallized and analyzed as discrete proteins. We have previously shown that these specific

formyltransferase and decarboxylase domains are catalytically functional and together (expressed in trans) can substitute for the ArnA protein (expressed as a single polypeptide) in the biosynthesis of L-Ara4N modified lipid A (25). Thus we do not expect any artifacts from studying the domains separately. All bacteria known to modify their lipid A molecules with L-Ara4N have full length ArnA orthologues. Although their lipid As have not been characterized, the *Raslstonia* spp. *solanacearum* and *metallidurans* genomes contain all of the genes required for the L-Ara4N modification of lipid A, and have the two domains of ArnA annotated as distinct proteins. Thus, separate ArnA formyltransferase and ArnA decarboxylase domains or enzymes may occur *in vivo* in some bacteria. Under the conditions we have employed, if an association between the discrete ArnA domains exists, it is very weak. Our data also suggest that ArnA does not interact strongly with ArnB.

The ArnA formyltransferase structure has two domains, both of which are structurally similar to the two domains of FMT, but only the N-terminal domain bears similarity to GART. The key catalytic residues, N102, H104 and D140 in ArnA, are found in the N-terminal domain and are conserved through all three proteins. This suggests that the mechanism of formyl transfer described by Almasy *et al* (26) also occurs in ArnA. For ArnA this transfer is likely to involve direct nucleophilic attack on the formyl carbon of *N*-10-fTHF by the 4'-amine of UDP-L-Ara4N, producing a tetrahedral intermediate. Similarly, D140 could be the ultimate base accepting the proton from the 4'-amine of UDP-L-Ara4N, although the proton may transfer first to H104. As in GART, N102 is positioned close to the formyl binding site and could stabilize the negatively charged oxygen of the intermediate (26) (Figure 3a).

It is interesting that the ArnA formyltransferase C-terminal domain is present in FMT, but absent in GART. The role of this domain in FMT is to bind a nucleotide (adenosine) in the tRNA substrate, to correctly orient this substrate at the active site (41). A similar role appears to be played in ArnA; this domain interacts with the nucleotide UMP (Figure 3a) presumably to appropriately orient the 4'-amine of the L-Ara4N moiety. The GART substrate, glycinamide ribonucleotide, is much smaller than the FMT substrate, L-methionyl-tRNA and closer in size to the sugar-nucleotide substrate for ArnA. Interestingly, the N-terminal domain of GART is sufficient to bind both substrate and cofactor. The C-terminal domain of the ArnA formyltransferase may also be important in the full length ArnA, acting to spatially separate the formyltransferase and decarboxylase domains to avoid interference with each other.

The overall structure of the ArnA decarboxylase domain confirms it belongs to the SDR class of enzymes, consistent with the first step in the reaction being hydride abstraction. Mutations remote from the catalytic triad and NAD^+ would not be expected to perturb hydride transfer. Yet mutation of either S433 or E434 (both of which are conserved) significantly impact catalysis but do not significantly affect UDP-GlcUA or NAD^+ binding (Table 2). Furthermore, the S433A and E434A mutants still produce the decarboxylated product, UDP-Ara4O, and do not accumulate the 4'-keto-UDP-GlcUA intermediate. From the biochemical data we draw two important conclusions.

First, the decarboxylation of 4'-keto-UDP-GlcUA is not spontaneous but is indeed enzyme catalyzed. The inactivity of E434Q and the over thirty fold reduced activity of S433A and E434A, suggest that the initial NAD^+ -dependent oxidation of the 4'-OH in the substrate UDP-GlcUA, has not been decoupled from the decarboxylation leading to the UDP-Ara4O product. Gatzee-Topalova *et al.* (27) postulated that E434 serves as a steric blocking group preventing rotation of the substrate and the 4'-keto intermediate in the active site. Implicit in this mechanism is that the rates of decarboxylation and hydride transfer are significantly slower than rotation of the carbohydrate at the active site. We felt that E434 may instead be

involved in catalysis and decided to probe its role by making the isosteric E434Q mutant. Although E434Q should be able to block rotation as efficiently as the native enzyme, this mutant is inactive. Further, the E434A mutant, which should have sufficient volume in the active site to allow ring rotation, retains activity. Together these data suggest that E434 serves a different purpose than blocking rotation the glucuronic acid moiety. E434 could act as a base ensuring deprotonation of the carboxylic acid, thus promoting decarboxylation upon formation of 4'-keto sugar (Figure 5). The observed activity of the E434A mutant may be accounted for by water molecules filling the newly created void to provide a bridge between the carboxylate of UDP-GlcUA and a second unidentified base in the protein, permitting the deprotonation of UDP-GlcUA. In E434Q the water bridge to the second base would not be present, and therefore decarboxylation does not occur. We suggest S433, by hydrogen bonding to the carboxylate group, may aid the decarboxylation. The S434A mutant shows that this serine does not have a significant role in substrate recognition (three fold change in apparent K_m).

The second conclusion is more subtle; in the step wise mechanism of decarboxylation, why should decarboxylation affect the preceding step of hydride transfer? The experimental evidence shows that these mutations do not perturb the structure nor do they significantly affect NAD^+ or UDP-GlcUA binding (where activity can be measured). One might therefore expect an accumulation of NADH as the enzyme generates the 4'-keto intermediate. That NADH production does not occur suggests that hydride transfer has an unfavorable free energy. Apparently, the equilibrium for the oxidative step lies far to the substrate (UDP-GlcUA) rather than the intermediate (4'-keto-UDP-GlcUA). It is the decarboxylation step, which is effectively irreversible, that serves to pull the substrate to product. When the irreversible step is either abolished or slowed significantly the hydride remains principally on the substrate, consistent with a slower rate of NAD^+ consumption.

Re-examining other carbohydrate utilizing SDR enzymes suggests that this may be a common theme. In RmlB, an irreversible step, the elimination of water, occurs after hydride transfer from substrate. Prior to that the reaction is reversible and both substrate and keto sugar have been shown to exist (45,46). GalE oxidizes and reduces its substrates without releasing the intermediate keto sugar (47,48). Thus, in each case, the equilibrium between oxidized substrate and NAD^+ is finely balanced. High resolution crystallographic studies and electronic structure calculations have shown that SDR enzymes fine tune the redox potential of $NAD(P)H$ by distortion of the nicotinamide ring (44). ArnA, GalE and RmlB operate on quite similar carbohydrates (Glc and GlcUA), and their active sites have similar volumes. By poising the equilibrium of hydride transfer towards substrate such that only a second irreversible step will create flux through the pathway, a subtle substrate specificity can be created by the second irreversible step. ArnA is unable to reduce NAD^+ in the presence of UDP-Glc (very similar to UDP-GlcUA) and therefore does not oxidize the C4' position. The structure of ArnA does not reveal any reason why UDP-Glc should not bind; it is slightly smaller than UDP-GlcUA. The presence of E434 and S433 in ArnA would prevent catalysis of the dehydration of UDP-Glc, since the proton source required for water elimination is absent. By ensuring in ArnA, that the equilibrium for the first oxidation reaction favors substrate, accumulation of reactive and undesired 4'-keto sugars closely related in structure to the true substrate are prevented. This may well be a general feature of these very similar SDR enzymes.

Acknowledgments

JHN is a BBSRC career development fellow. GJW and JHN are supported by the Wellcome Trust. CRHR and SDB are supported by National Institutes of Health grant GM-51310.

Abbreviations

CAMP	cationic antimicrobial peptide
FMT	<i>N</i> -10-formyltetrahydrofolate:L-methionyl-tRNA <i>N</i> -formyltransferase
GaIE	UDP-galactose-4-epimerase
GART	glycinamide ribonucleotide transformylase
Glc	glucose
GlcUA	glucuronic acid
HPLC	high pressure liquid chromatography
L-Ara4N	4-amino-4-deoxy-L-arabinose
LPS	lipopolysaccharide
NAD⁺	nicotinamide adenosine dinucleotide
NADP⁺	nicotinamide adenosine dinucleotide phosphate
<i>N</i>-5-fTHF	<i>N</i> -5-formyltetrahydrofolate
<i>N</i>-10-fTHF	<i>N</i> -10-formyltetrahydrofolate
RmlB	dTDP-glucose-4,6-dehydratase
Rmsd	root mean square deviation
SDR	short chain dehydrogenase
Se-Met	selenomethionine
dTDP	thymidine diphosphate
UDP	uridine diphosphate
UMP	uridine monophosphate

References

1. Raetz CRH. *Annu Rev Biochem.* 1990; 59:129–170. [PubMed: 1695830]
2. Medzhitov R, Janeway C Jr. *N Engl J Med.* 2000; 343:338–344. [PubMed: 10922424]
3. Dinarello CA. *Blood.* 1991; 77:1627–1652. [PubMed: 1826616]
4. Beutler B, Cerami A. *Annu Rev Biochem.* 1988; 57:505–518. [PubMed: 3052281]
5. Ohl ME, Miller SI. *Annu Rev Med.* 2001; 52:259–274. [PubMed: 11160778]
6. Parrillo JE. *N Engl J Med.* 1993; 328:1471–1477. [PubMed: 8479467]
7. Zasloff M. *Nature.* 2002; 415:389–395. [PubMed: 11807545]
8. Hoffmann JA, Kafatos FC, Janeway CA, Ezekowitz RA. *Science.* 1999; 284:1313–1318. [PubMed: 10334979]
9. Scott MG, Hancock RE. *Crit Rev Immunol.* 2000; 20:407–431. [PubMed: 11145218]
10. Matsuzaki K. *Biochim Biophys Acta.* 1999; 1462:1–10. [PubMed: 10590299]
11. Shai Y. *Biochim Biophys Acta.* 1999; 1462:55–70. [PubMed: 10590302]
12. Gunn JS, Ryan SS, Van Velkinburgh JC, Ernst RK, Miller SI. *Infect Immun.* 2000; 68:6139–6146. [PubMed: 11035717]
13. Gunn JS. *J Endotoxin Res.* 2001; 7:57–62. [PubMed: 11521084]
14. Zhou Z, Lin S, Cotter RJ, Raetz CRH. *J Biol Chem.* 1999; 274:18503–18514. [PubMed: 10373459]
15. Wosten MM, Kox LF, Chamnongpol S, Soncini FC, Groisman EA. *Cell.* 2000; 103:113–125. [PubMed: 11051552]

16. Nummila K, Kilpelainen I, Zahringer U, Vaara M, Helander IM. *Mol Microbiol.* 1995; 16:271–278. [PubMed: 7565089]
17. Zhou Z, Ribeiro AA, Lin S, Cotter RJ, Miller SI, Raetz CRH. *J Biol Chem.* 2001; 276:43111–43121. [PubMed: 11535603]
18. Gunn JS, Lim KB, Krueger J, Kim K, Guo L, Hackett M, Miller SI. *Mol Microbiol.* 1998; 27:1171–1182. [PubMed: 9570402]
19. Guo L, Lim KB, Gunn JS, Bainbridge B, Darveau RP, Hackett M, Miller SI. *Science.* 1997; 276:250–253. [PubMed: 9092473]
20. Baker SJ, Gunn JS, Morona R. *Microbiology.* 1999; 145(Pt 2):367–378. [PubMed: 10075419]
21. Breazeale SD, Ribeiro AA, Raetz CRH. *J Biol Chem.* 2002; 277:2886–2896. [PubMed: 11706007]
22. Breazeale SD, Ribeiro AA, Raetz CRH. *J Biol Chem.* 2003; 278:24731–24739. [PubMed: 12704196]
23. Trent MS, Ribeiro AA, Lin S, Cotter RJ, Raetz CRH. *J Biol Chem.* 2001; 276:43122–43131. [PubMed: 11535604]
24. Noland BW, Newman JM, Hendle J, Badger J, Christopher JA, Tresser J, Buchanan MD, Wright TA, Rutter ME, Sanderson WE, Muller-Dieckmann HJ, Gajiwala KS, Buchanan SG. *Structure.* 2002; 10:1569–1580. [PubMed: 12429098]
25. Breazeale SD, Ribeiro AA, McClerren AL, Raetz CRH. *J Biol Chem Epub ahead of publications.* 2005
26. Almasy RJ, Janson CA, Kan CC, Hostomska Z. *Proc Natl Acad Sci U S A.* 1992; 89:6114–6118. [PubMed: 1631098]
27. Gatzeva-Topalova PZ, May AP, Sousa MC. *Biochemistry.* 2004; 43:13370–13379. [PubMed: 15491143]
28. Beardsley GP, Moroson BA, Taylor EC, Moran RG. *J Biol Chem.* 1989; 264:328–333. [PubMed: 2909524]
29. Double S. *Macromolecular Crystallography.* 1997; (Pt A 276):523–530.
30. Leslie AGW. *Newsletter on Protein Crystallography.* 1992; 26:1–10.
31. Evans PR. *Newsletter on Protein Crystallography.* 1997; 33:22–24.
32. Terwilliger TC, Berendzen J. *Acta Crystallogr r.* 1999; D55:849–861.
33. Terwilliger TC. *Acta Crystallogr.* 2000; D56:965–972.
34. Cowtan K. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography.* 1994; 31:34–38.
35. Collaborative Computational Project, N. *Acta Crystallogr r.* 1994; D50:760–763.
36. Perrakis A, Morris R, Lamzin VS. *Nat Struct Biol.* 1999; 6:458–463. [PubMed: 10331874]
37. Jones TA, Zou J-Y, Cowan SW, Kjeldgaard M. *Acta Crystallographica Section A.* 1991; 47:110–119.
38. Emsley P, Cowtan K. *Acta Crystallogr y.* 2004; D60:2126–2132.
39. Murshudov GN, Vagin AA, Lebedev A, Wilson KS, Dodson EJ. *Acta Crystallogr.* 1999; D55:247–255.
40. Schuttelkopf AW, van Aalten DMF. *Acta Crystallogr.* 2004; D60:1355–1363.
41. Schmitt E, Panvert M, Blanquet S, Mechulam Y. *EMBO J.* 1998; 17:6819–6826. [PubMed: 9843487]
42. Allard ST, Beis K, Giraud MF, Hegeman AD, Gross JW, Wilmouth RC, Whitfield C, Graninger M, Messner P, Allen AG, Maskell DJ, Naismith JH. *Structure.* 2002; 10:81–92. [PubMed: 11796113]
43. Thoden JB, Hegeman AD, Wesenberg G, Chapeau MC, Frey PA, Holden HM. *Biochemistry.* 1997; 36:6294–6304. [PubMed: 9174344]
44. Beis K, Allard ST, Hegeman AD, Murshudov G, Philp D, Naismith JH. *J Am Chem Soc.* 2003; 125:11872–11878. [PubMed: 14505409]
45. Hegeman AD, Gross JW, Frey PA. *Biochemistry.* 2002; 41:2797–2804. [PubMed: 11851427]
46. Hegeman AD, Gross JW, Frey FA. *Biochemistry.* 2001; 40:6598–6610. [PubMed: 11380254]
47. Thoden JB, Frey PA, Holden HM. *Biochemistry.* 1996; 35:2557–2566. [PubMed: 8611559]
48. Thoden JB, Frey PA, Holden HM. *Biochemistry.* 1996; 35:5137–5144. [PubMed: 8611497]

49. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. *Journal of Applied Crystallography*. 1993; 26:548–558.

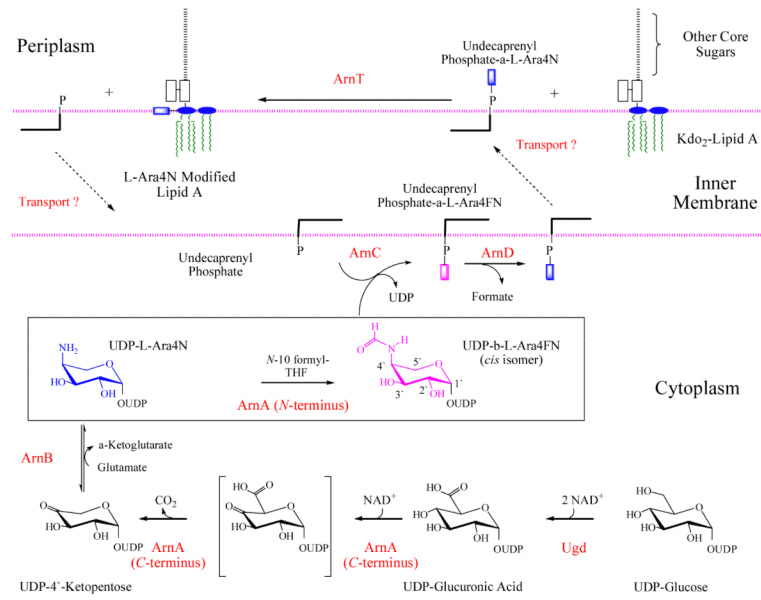
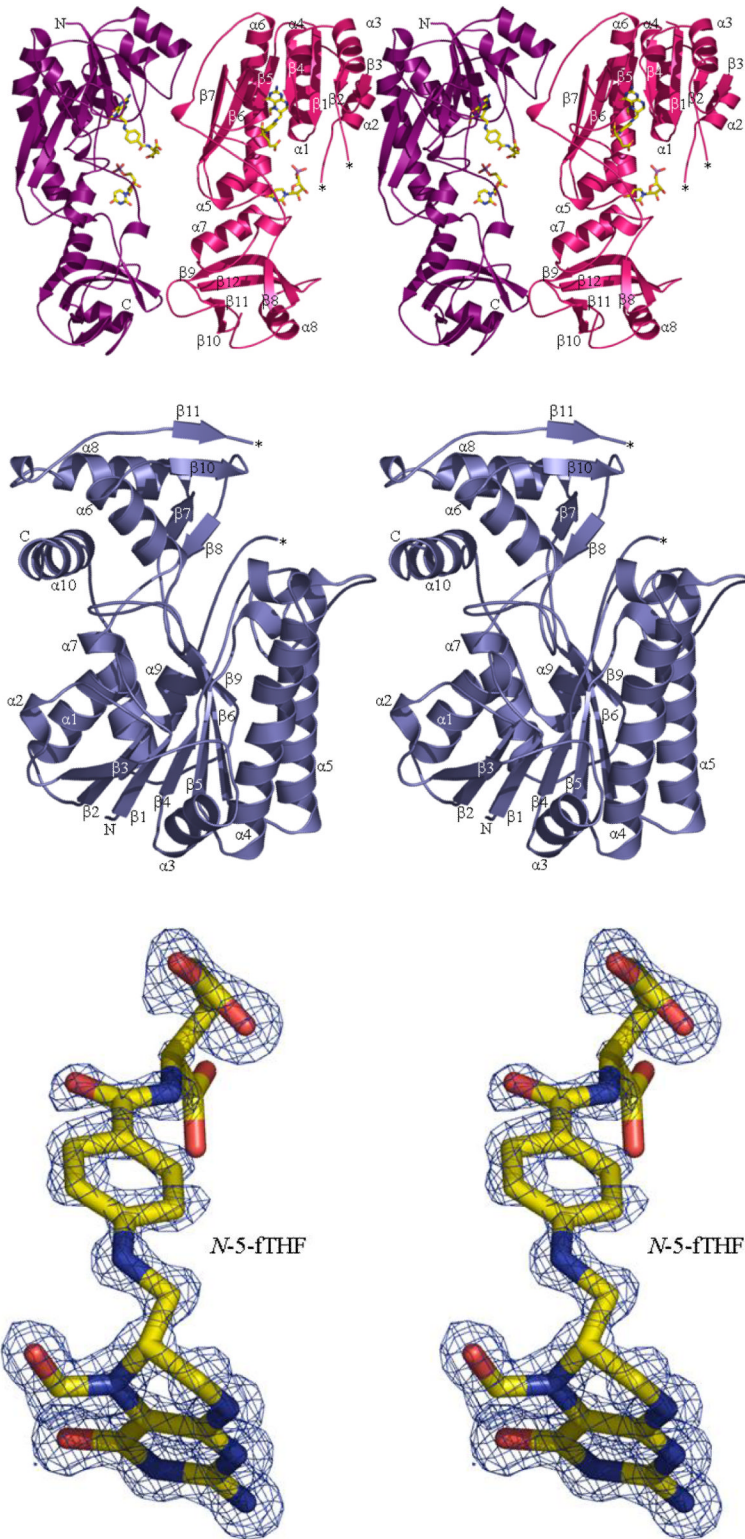


Figure 1.
The assembly process for the modification of LPS by L-Ara4N (22,24).



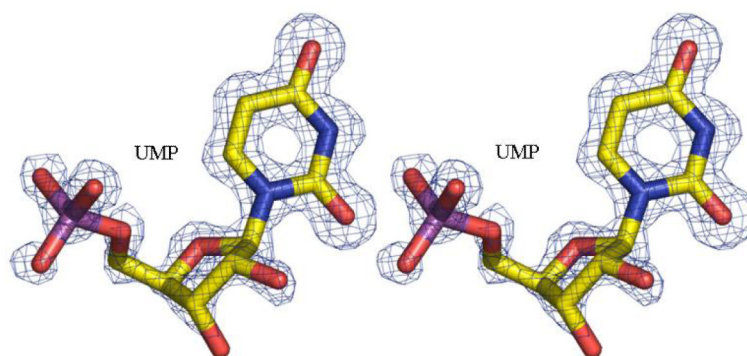


Figure 2.

the structure of ArnA

2a The structure of the formyltransferase domain. The protein is represented as ribbons and the secondary structure elements are numbered as the text. A loop (residues N35-A40) that is missing in the structure is labeled *. The two subdomains are visible, the N-terminal subdomain is at the top of the figure and extends to the uridine ring, the C-terminal subdomain sits below the ring. The dimer found in the crystal is shown. The decarboxylase domain is attached to the C-terminus (labeled C).

2b The structure of the decarboxylase domain. The protein is represented as ribbons and the secondary structure elements are numbered. A loop (residues V604-D615) that is missing in the structure is labeled *. The monomeric unit is shown. The N-terminus of this domain (labeled N) is attached to the formyltransferase domain.

2c The unbiased Fo-Fc electron density map for *N*-5-fTHF found in the formyltransferase domain. The molecules are shown in stick with atoms colored; carbon yellow, nitrogen blue, oxygen red and phosphorous purple. The map is contoured at 3σ ($0.22\text{e}\text{\AA}^{-3}$)

2d The unbiased Fo-Fc electron density map for UMP found in the formyltransferase domain. Contouring and atomic color are the same as Figure 2c.

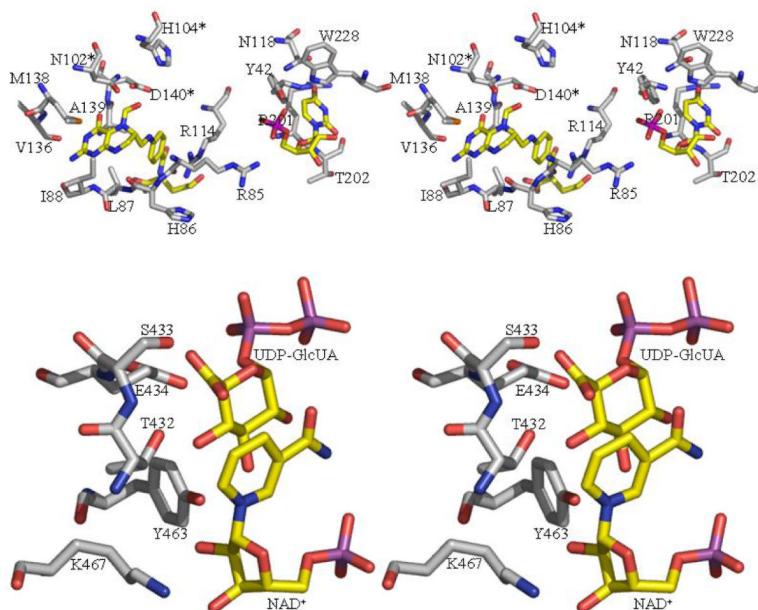


Figure 3.

The active sites of ArnA

3a The active site of formyl transferase. The key catalytic residues are shown (*) as are the two substrate mimics (UMP and *N*-5-fTHF) located experimentally and those residues which recognize them. The atomic color scheme is the same as Figure 2c

3b The active site of the decarboxylase enzyme. We have modeled in the UDP-GlcUA substrate to generate a conceptual model. The model is based on a superposition of the RmlB substrate complex (42). The catalytic triad residues are shown as are the residues we predict to be important in the decarboxylation step (S433 and E434). These residues superimpose with D135 and E136 of RmlB which are responsible for catalysis of the dehydration step in RmlB transformation (42). The atomic color scheme is the same as Figure 2c.


```

1                                     41#
ArnA ...MKTVVF AYHDMGCLGI EALLAAGYEI SAIFTHTDNP ...GEKAFYF
FMT MSESRLRIIFA GTPDFAARHL DALLSSGHNV VGVFTQPDPR AGRGKKLMPS
GART ..MNIVVLIS GNGSNLQAI DACKTNKIKG TVRAVFSNKA DAFGLERA.R
      +                                     +

44                                     ### 93
ArnA SVARLAAERG IPVYAPDNVN HPLWVERIAQ LSPDVIFSFY YRHLYDEIL
FMT PVKVLAAEKG LPVFQPVSLR PQENQQLVAE LQADVMMVVA YGLLIPKAVL
GART QAGIATHTLI ASAFDSREAY DRELIHEIDM YAPDVVVLG FMRILSPAFV
      ++

94      * *      ‡ #      †† *143
ArnA QLAPAGAFNL HGSLLPKYRG RAPLNWVLVN GETETGVTLH RMVKRADAGA
FMT EMPRLGCINV HGSLLPRWRG AAPIQRSLWA GDAETGVTIM QMDVGLDTGD
GART SHYAGRLLNI HPSLLPKYFG LHTHRQALEN GDEEHGTSVH FVTDELGGP
      +++++ +      + + + +      +

144                                     193
ArnA IVAQLRIAIA PDDIAITLHH KLCHAARQLL EQTLPAIKHG NILEIAQREN
FMT MLYKLSCPIT AEDTSGTLYD KLAELGPQGL ITTLKQLADG TAKPEVQDET
GART VILQAKVPVF AGDSEDDITA RVQTQEHAIV PLVISWFADG RLK...MHEN
      +                                     +

194      ##      #                                     243
ArnA EATCFGR RTP DDSFLEWHKP ASVLHNMVRA VADPWPGAFS YVGNQKFTVW
FMT LVTYAEKLSK EEARIDWSLS AAQLERCIRA F.NPWPMSWL EIEGQPVKVV
GART AAWLDGQRLP PQGYAADE...

244                                     291
ArnA SSRVHPHASK AQP GSVISVA P..LLIACGD GALEIVTGQA GDGITMQGSQ
FMT KASVIDTATN AAPGTILEAN KQGIQVATGD GILNLLSLQP AGKKAMSAQD
GART .....

292                                     305
ArnA LAQTLG..LV QGSRLN
FMT LLNSRREWFV PGNRLV
GART .....

```

Figure 4. Sequence alignment of formyl transfer domains, numbering corresponds to full length ArnA. The key conserved catalytic residues are marked * as in Figure 3a, other conserved residues have a + underneath. ArnA residues that recognize *N*-5-fTHF and UMP are marked with ‡ and # respectively.

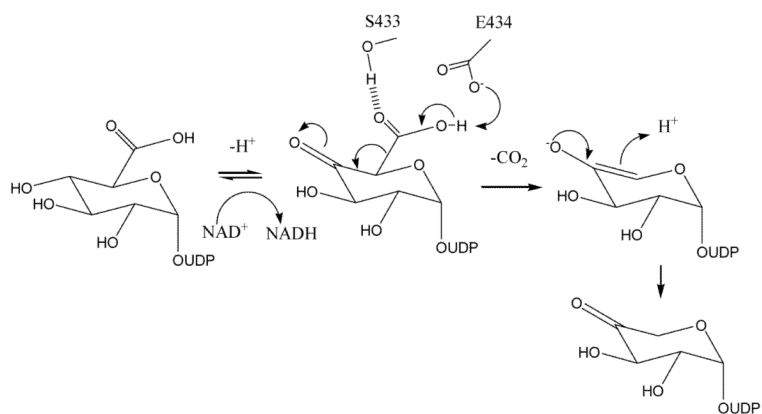


Figure 5.
A plausible mechanism for the catalysis of the decarboxylation reaction.

Table 1
Crystallographic statistics for data collected on native and selenomethionine Arna decarboxylase and formyltransferase domains

Data collection	Arna decarboxylase domain				Arna formyltransferase domain			
	Peak	Inflection	Remote	Native	Peak	Inflection	Remote	Native
Resolution (Å)	61-3.35				24-1.6			
Highest shell (Å)	3.43-3.35				1.63-1.6			
Wavelength (Å)	0.9798	0.9710	0.9392	88-2.30	0.9793	0.9796	0.9393	1.26-1.2
Unique reflections	15066	15065	15047	0.933	77800	77914	77900	0.934
Multiplicity	65 (67)	21 (21)	21 (21)	25245	4.4 (4.5)	4.4 (4.5)	4.3 (4.4)	4.9 (4.9)
Completeness (%)	100 (100)	100 (100)	100 (100)	97.8 (100)	96.8 (95.8)	96.9 (95.8)	96.8 (95.8)	98.3 (98.3)
Rmerge (%)	14.3 (31.1)	12.4 (27.1)	13.4 (29.7)	7.0 (36.9)	5.1 (11.5)	5.1 (11.3)	5.1 (11.3)	6.2 (29.7)
I/σ	4.4 (2.2)	4.7 (1.5)	5.3 (2.5)	7.8 (2.1)	8.8 (6.3)	8.9 (6.2)	8.8 (6.2)	7 (2.4)
Structure solution								
Monomers in a.u.	1				2			
Se sites	4				8			
z-score	13.8				21.3			
Phasing power	0.52				0.64			
Refinement								
Rmsd ² bond length (Å)/angles (°)	0.019/1.63				0.019/1.93			
Rfactor/R _{FREE} (%)	18.6/23.0 (21.1/26.5)				13.5/15.7 (14.4/18.2)			
Residues in most favored regions ³ (%)	90				92			
Residues in allowed regions ³ (%)	10				8			

	ArnA decarboxylase domain				ArnA formyltransferase domain			
Data collection								
PDB code	Peak	Inflection	Remote	Native	Peak	Inflection	Remote	Native
				2b11				2bln

¹Numbers in parenthesis correspond to highest resolution shell.

²Rmsd root mean square deviation

³These refer to the Ramachandran plot and are defined by PROCHECK (49).

Table 2

Enzyme assay results

Enzyme	Specific activity ($\mu\text{M min}^{-1}\text{mg}^{-1}$)	UDP-GlcUA		NAD+		Protein (μM)
		V_{max} ($\mu\text{M min}^{-1}$)	Apparent K_m (mM)	V_{max} ($\mu\text{M min}^{-1}$)	Apparent K_m (mM)	
AraA Decarboxylase domain	350	6.5	0.7	6.8	1.3	0.5
S433A Decarboxylase mutant	8.8	0.7	0.2	0.8	1.4	2.5
E434A Decarboxylase mutant	3.3	0.3	0.4	0.3	1.6	2.5
E434Q Decarboxylase mutant	No Activity	N/A	N/A	N/A	N/a	2.5