Crystal Structure of dTDP-4-keto-6-deoxy-D-hexulose 3,5-Epimerase from Methanobacterium thermoautotrophicum Composed with dTDP*

Received for publication, April 9, 2000, and in revised form, May 22, 2000
Published, JBC Papers in Press, May 25, 2000, DOI 10.1074/jbc.C000238200

Dinesh Christendat‡§§, Vivian Saridakis§§, Akil Dharamsi§§§, Alexei Bochkarev‡‡, Emil F. Pai§§, Cheryl H. Arrowsmith§§§ §§§, and Aled M. Edwards‡§§§§§

From the ‡Division of Molecular and Structural Biology, Ontario Cancer Institute, Toronto, Ontario M5G 2M9, Canada, the §Department of Medical Biophysics, ‡Baniting and Best Department of Medical Research, University of Toronto, Toronto, Ontario M5S 1L6, Canada, **Integrated Proteomics Inc., Toronto, Ontario M5G 2M9, Canada, and the §§Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, BRC-466, Oklahoma City, Oklahoma 73190

Deoxythymidine diphosphate (dTDP)-4-keto-6-deoxy-D-hexulose 3,5-epimerase (RmlC) is involved in the biosynthesis of dTDP-L-rhamnose, which is an essential component of the bacterial cell wall. The crystal structure of RmlC from Methanobacterium thermoautotrophicum was determined in the presence and absence of dTDP, a substrate analogue. RmlC is a homodimer comprising a central jelly roll motif, which extends in two directions into longer β-sheets. Binding of dTDP is stabilized by ionic interactions to the phosphate group and by a combination of ionic and hydrophobic interactions with the base. The active site, which is located in the center of the jelly roll, is formed by residues that are conserved in all known RmlC sequence homologues. The conservation of the active site residues suggests that the mechanism of action is also conserved and that the RmlC structure may be useful in guiding the design of antibacterial drugs.

Proteins whose expression and activity are restricted to prokaryotes are attractive antibiotic targets. The comparative analysis of comprehensive genome data bases has uncovered a large set of such proteins, which includes enzymes involved in bacterial-specific intermediary metabolism and those involved in the biosynthesis of the bacterial cell wall. The bacterial cell wall comprises a large number of carbohydrates that are not found in mammalian cells, one of which is the activated form of L-rhamnose, dTDP-L-rhamnose. dTDP-L-rhamnose is synthesized from α-D-glucose 1-phosphate by a set of four bacterial-specific enzymes, called RmlA through D, whose sequences are highly conserved between different organisms. RmlA, glucose-1-phosphate thymidylyltransferase, catalyzes the synthesis of dTDP-D-glucose from dTTP and α-D-glucose 1-phosphate. The next enzyme in the pathway, dTDP-1-D-glucose 4,6-dehydratase (RmlB) reduces dTDP-D-glucose to dTDP-4-keto-6-deoxy-D-glucose in an NADH-dependent reaction. RmlC, dTDP-4-keto-6-deoxy-D-hexulose 3,5-epimerase, then converts dTDP-4-keto-6-deoxy-D-glucose to dTDP-4-keto-L-rhamnose. Finally, RmlD, dTDP-4-keto-L-rhamnose reductase, reduces dTDP-4-keto-L-rhamnose to dTDP-L-rhamnose in an NADPH-dependent reaction (1, 2).

The enzymatic mechanism of dTDP-L-rhamnose biosynthesis began to be elucidated more than 30 years ago. More recent studies have focused on the molecular genetics and structural biology of the corresponding enzymes (1, 2). In this study, we report the crystal structures of the apo and a ligand-bound form of the RmlC homologue from Methanobacterium thermoautotrophicum, an organism that is one of the target organisms in our structural proteomics effort. Structural analysis of RmlC has uncovered significant structural homology to concanavalin A and has allowed us to hypothesize a mechanism for the dTDP-4-keto-6-deoxy-D-glucose epimerization reaction.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—The RmlC gene from M. thermoautotrophicum genomic DNA was amplified by polymerase chain reaction and cloned into the pET15β expression vector (Novagen). Recombinant dTDP-4-keto-6-deoxy-D-hexulose epimerase (RmlC) was expressed in Escherichia coli BL21 Gold (DE3) cells (Strategene) harboring a plasmid encoding three rare E. coli tRNA genes (AGG and AGA for Arg and ATA for Ile). Conditions for protein expression and purification were similar to those in the Qiagen protein purification handbook except that a heat step (55 °C for 10 min) and a centrifugation step were introduced after cell lysis to remove most contaminating E. coli proteins. Purified RmlC was dialyzed against 10 mM HEPES and 500 mM NaCl and concentrated to 10 mg/ml using BioMax concentrators (Millipore). For the preparation of selenomethionine (Se-Met) protein, RmlC was expressed in a methionine auxotroph strain B834(DE3) (Novagen) and purified under the same conditions as native RmlC with the addition of 5 mM 4-mercaptoethanol in all buffers.

Gel Filtration—Gel filtration of RmlC was performed with a Superdex 200 prep 16/60 (Amersham Pharmacia Biotech) column equili-
Crystal Structure of RmlC with dTDP

Numbers in parentheses represent values in the highest resolution shell (native 1.55–1.50 Å; SeMet 2.07–2.00 Å; dTDP 1.81–1.75 Å). R \text{sym} = \Sigma I - (I)/\Sigma I, where I is the observed integrated intensity, \langle I \rangle is the average integrated intensity obtained from multiple measurements, and the summation is over all observed reflections.

| X-ray diffraction data | Native | SeMet (peak) | SeMet (edge) | SeMet (remote) | dTDP |
|------------------------|--------|--------------|--------------|----------------|------|
| Space group            | C2     | C2           | C2           | C2             | C2   |
| Unit cell (Å³)         | 67.7 × 53.1 × 51.7 | 67.7 × 53.5 × 51.7 | 67.7 × 53.5 × 51.7 | 67.7 × 53.5 × 51.7 | 67.8 × 53.4 × 51.5 |
| β = 96.6°              | β = 96.6° | β = 96.6°   | β = 96.6°    | β = 96.6°      | β = 97.2° |
| Resolution (Å)         | 30–1.5 | 30–2.0       | 30–2.0       | 30–2.0         | 30–1.75 |
| Wavelength (Å)         | 1.000  | 0.9793       | 0.9796       | 0.9537         | 1.5417 |
| No. of unique reflections | 180,757 | 45,252      | 45,363       | 45,894         | 63,707 |
| No. of observations    | 53.1   | 53.5         | 53.5         | 53.1           | 53.4 |
| No. of protein atoms   | 1518   | 1518         | 1518         | 1518           | 1518  |
| No. of water molecules | 127    | 119          | 119          | 119            | 119   |
| No. of dTDP atoms      | 25     | 25           | 25           | 25             | 25    |
| Root mean square       | 0.015  | 0.005        | 0.005        | 0.005          | 0.005 |
| Bond lengths (Å)       | 1.80   | 1.30         | 1.30         | 1.30           | 1.30  |
| Improper angles (°)    | 25.70  | 25.70        | 25.70        | 25.70          | 25.70 |
| Average B-factor (Å²)  | 13.70  | 18.50        | 18.50        | 18.50          | 18.50 |

FIG. 1. A, ribbon diagram of an RmlC subunit with a ball-and-stick model of complexed dTDP. The jelly roll structural motif is shown by the green and red β-strands. The secondary structure elements are labeled as depicted in the text. This figure was prepared using Molscript (13) and Raster3D (14). B, stereo view of the Cα trace of a subunit of RmlC. The numbers refer to the amino acid residues.

TABLE I
Summary of data collection statistics

|                     | Native | SeMet (peak) | SeMet (edge) | SeMet (remote) | dTDP |
|---------------------|--------|--------------|--------------|----------------|------|
| Completeness (%)    | 99.7   | 99.9         | 99.9         | 100            | 99.3 |
| Rfree (%)           | 18.3   | 19.6         | 19.6         | 19.6           | 21.1 |
| No. of observations | 180,757 | 45,252      | 45,363       | 45,894         | 63,707 |
| No. of protein atoms| 1518   | 1518         | 1518         | 1518           | 1518  |
| No. of water molecules | 127    | 119          | 119          | 119            | 119   |
| No. of dTDP atoms   | 25     | 25           | 25           | 25             | 25    |
| Root mean square    | 0.015  | 0.005        | 0.005        | 0.005          | 0.005 |
| Bond lengths (Å)    | 1.80   | 1.30         | 1.30         | 1.30           | 1.30  |
| Improper angles (°) | 25.70  | 25.70        | 25.70        | 25.70          | 25.70 |
| Average B-factor (Å²)| 13.70  | 18.50        | 18.50        | 18.50          | 18.50 |

FIG. 2. Overview of the dimeric structure of RmlC. Ribbon diagram of the RmlC dimer with a ball-and-stick model of complexed dTDP. Each subunit is colored differently. A, 2-fold axis of symmetry in the plane. B, rotation of A by 90° in the plane.

Crystallography was performed at 4 °C at a flow rate of 0.5 ml/min.

Crystallization—An initial crystallization condition was obtained with a sparse crystallization matrix (Hampton Research Crystal ScreenTM I) using the hanging drop vapor diffusion technique. This condition was modified slightly by varying the pH and concentration of polyethylene glycol and yielded crystals suitable for native and MAD data collection. The best crystals grew in 10% polyethylene glycol 4000 and yielded crystals suitable for native and MAD data collection. The best crystals grew in 10% polyethylene glycol 4000 and 100 mM sodium acetate at pH 4.6 in 2–4 days at 22 °C using hanging drops (3 μl:3 μl protein:precipitant ratio). They reached approximate dimensions of 600 × 200 × 200 microns². These crystals belonged to space group C2 with unit cell dimensions 67.7 Å × 53.1 Å × 51.7 Å and β = 96.6°. There was a single molecule in the asymmetric unit and the Matthews coefficient was 2.3 Å³/dalton resulting in an estimated solvent content of 46%. Soaking of RmlC crystals was carried out in 10 mM dTDP with 10% polyethylene glycol 4000 and 100 mM sodium acetate at pH 4.6 for 4 h.

X-ray Diffraction and Structure Determination—The structure of RmlC was determined by the MAD method using selenium as the anomalous scatterer. A three-wavelength MAD experiment was performed at the BioCARS 14BMD beamline at the Advanced Photon Source. The high resolution data of the native crystal were also collected with the BioCARS 14BMD beamline. The MAD and native data were processed and scaled with the DENZO/SCALEPACK (3) suite of programs. Three selenium sites were located using SOLVE (4) and refined using PHASES (5). Solvent flattening was done using PHASES. Model building was done with O (6). Crystallography and NMR system (7) was used for refinement with multiple rounds of minimization, simulated annealing, B-group, and individual B-factor refinement fol-
followed by manual rebuilding. Most of the water molecules were picked using crystallography and NMR system and additional ones were manually added after manual verification using O. The water molecules were picked using the following criteria in O: a peak of at least 2.5 $\sigma$ on an $F_o - F_c$ map, a peak of at least 1.0 $\sigma$ on a $2F_o - F_c$ map, and reasonable intermolecular interactions. The crystallographic data collection and refinement statistics are given in Tables I and II, respectively.

RESULTS AND DISCUSSION

Structure Determination—The structure of selenomethionine-enriched RmlC was determined by the MAD method and refined against 1.5 Å resolution data to a working $R$-factor of 0.183 and a free $R$-factor of 0.211. The refined apo model contains 183 amino acids (residues 3–185) and 127 water molecules (Fig. 1). The electron density of the apo form, which was used to build the model, is of excellent quality except for the loop between residues 140 and 144. The dTDP complex model was refined against 1.75 Å resolution data to a working $R$-factor of 0.195 and a free $R$-factor of 0.224. This model contains 183 amino acid residues, 119 water molecules, and one molecule of dTDP (Fig. 1). The first two amino acids at the N terminal are not visible in the electron density map in either model. PROCHECK (8) was used to evaluate the stereochemistry of both of the refined models, which showed that more than 90% of the residues are in the allowed region and only one
amino acid (Glu-68) was in the disallowed regions, because it is present in a γ turn between β6 and β7.

Overview of the Structure—RmlC is a homodimer; this was confirmed by gel filtration analysis (data not shown). The monomer comprises thirteen β-strands and three short α-helices (Fig. 1). Eight of the β-strands are arranged in a central eight-stranded antiparallel β-sheet (strands β5A to β12A) that resembles a jelly roll (Fig. 1). Four other strands β1A, β2A, β3B, and β4B (from subunits A and B) extend from strands β5A, β7A, β10A, and β11A from the jelly roll to form an eight-stranded anti-parallel β-sheet. A second β-sheet is formed by β13A aligned in an antiparallel manner with strands β6A, β8A, β9A, and β11A. The helices are located on the periphery of the molecule. Helix 1 packs against strand β1 from the N-terminal β-sheet. Helices 2 and 3 flank the carboxyl terminus of the subunit and are also involved in important crystal packing interactions. Helix 2 also contributes to the active site of the same subunit.

The dimer interface is formed by an extensive set of hydrophobic and electrostatic contacts between β3 and β5, β7 and β7, and α1 and β5. Some of these ionic interactions include Arg-61 to Asp-24 via a water molecule and the formation of two salt bridges (Glu-52 to Arg-76 and Asp-50 to Lys-134). Hydrophobic interactions occur between residues Phe-33 Ala-36, Tyr-28, Arg-26 (aliphatic side chain), Val-48, Val-59, Ile-78, and Leu-138 at the subunit interface. These interactions result in a total buried surface area of 3,042 Å2 out of a total of 16,306 Å2 for the dimer.

A search for structural homologues using the program DALI (9) revealed that RmlC is homologous to concanavalin A, phosphomannose isomerase, and arabinose operon regulatory protein (AraC). The nearest structural neighbor is concanavalin A, which has a Z-score of 6.4 and root mean square deviation (r.m.s.d.) of 1.8 Å over 87 out of 178 Cα atoms. The overall core topology of all these molecules is similar to the jelly roll structural motif.

Location of the Active Site—Residues involved in substrate binding and catalysis were identified by determining the structure of RmlC in the presence of a substrate analogue, dTDP. The electron density map of the complex revealed a well-ordered dTDP with high occupancy (Fig. 3). The substrate-binding site is located in the center of a cavity formed by the jelly roll structural motif (which is at the middle of one face of one subunit) (Fig. 2). Residues from β-strands 3 and 4 from one subunit combine with β-strands 5, 6, 11, and 12 from the other subunit to form a complete active site. The active site is open at the center of each subunit to permit entry and exit of the ligand through the B-face (Fig. 2). The active site is lined with a number of charged residues (Gln-49, Asp-84, Asp-144, Asp-172, Glu-31, Lys-73, Lys-171, Glu-52, Arg-26, Arg-61, His-64, His-120, and Cys-135) and a number of residues with hydrogen-bonding potentials (Ser-53, Ser-55, Ser-169, Gln-49, Glu-3 and Asn-51), which together constitute the catalytic triad. The active site is also lined with aromatic residues (Trp-175, Phe-29, Phe-122, Tyr-133 and Tyr-139), which provide favorable environments for the base moiety of dTDP and potentially for the sugar moiety of the substrate (Fig. 4).

Comparison between Apo- and dTDP-bound dTDP-4-keto-6-deoxy-6-hexulose Epimerase—The structure of a subunit of the apo form of RmlC is very similar to that of the dTDP-bound
enzyme with an overall r.m.s.d. of 0.33 Å for 183 Cα atoms. There are, however, some notable differences between the apo- and dTDP-enzymes. The most prominent differences occur within residues 140–144, which are visible in the presence of dTDP. In the presence of the ligand, this loop becomes ordered, closing off a portion of the active site. This loop may be important in regulating the passage of the substrate/product into and out of the active site and may serve to keep the external solvent molecules away from the active site.

Substrate Binding—The dTDP portion of dTDP-4-keto-6-deoxy-α-hexulose anchors the substrate in the active site of the enzyme. dTDP binds between strands β5, β6, β11, and β12 of one subunit and β3 and β4 of the other subunit. Aromatic stacking is observed between Tyr-139 and Phe-29 and the base moiety of dTDP. In fact, the electron density of the side chains of Tyr-133, Tyr-139, and Lys-171 was observed only in the presence of dTDP. In the absence of dTDP, Tyr-139 stacks against the base moiety of dTDP and Lys-171 makes ionic interactions with an oxygen of the β-phosphate of dTDP through a water molecule. The base of dTDP is bound in an anticonformation relative to the ribose ring (Fig. 4) by hydrogen bonding to Glu-31B and Gln-49A. The diphosphate portion of dTDP is securely anchored to the protein by ionic interactions between the oxygens of the phosphates with Arg-61A and Arg-26B. In addition to these interactions, there are also a number of interactions between the phosphate oxygens and the enzyme, which are mediated by water molecules (waters 1035, 1036, 1071, and 1095).

Model for Enzymatic Mechanism—The use of three-dimensional structural information to generate hypotheses about reaction mechanisms and protein function is likely to be a common occurrence in structural genomics projects, which will provide structural information often in the absence of the corresponding biochemical information. In this instance, a possible reactive center(s) for the epimerization of hexulose by RmlC was determined by analyzing the three-dimensional structure of one subunit and b subunit. Aromatic stacking is observed between Tyr-139 and Phe-29 and the base moiety of dTDP.

Acknowledgments—We thank the staff of BioCARS for their help during data collection at Sector 14 of the Advanced Photon Source. We thank Ashleigh Tuite for help during crystallization, Steven Beasley for help with protein purification, and Matthew Kimber for helpful discussions.

Addendum—While this manuscript was under review, the structure of RmlC from Salmonella typhimurium was published (16).