Published in final edited form as:

Environ Microbiol. 2012 November ; 14(11): 2920–2934. doi:10.1111/j.1462-2920.2012.02856.x.

Tagaturonate–fructuronate epimerase UxaE, a novel enzyme in the hexuronate catabolic network in *Thermotoga maritima*

Irina A. Rodionova¹, David A. Scott¹, Nick V. Grishin², Andrei L. Osterman¹, and Dmitry A. Rodionov^{1,3,*}

¹Sanford-Burnham Medical Research Institute, La Jolla, CA 92037, USA

²University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

³Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow 127994, Russia

Summary

Thermotoga maritima is a marine hyperthermophilic microorganism that degrades a wide range of simple and complex carbohydrates including pectin and produces fermentative hydrogen at high yield. Galacturonate and glucuronate, two abundant hexuronic acids in pectin and xylan, respectively, are catabolized via committed metabolic pathways to supply carbon and energy for a variety of microorganisms. By a combination of bioinformatics and experimental techniques we identified a novel enzyme family (named UxaE) catalysing a previously unknown reaction in the hexuronic acid catabolic pathway, epimerization of tagaturonate to fructuronate. The enzymatic activity of the purified recombinant tagaturonate epimerase from *T. maritima* was directly confirmed and kinetically characterized. Its function was also confirmed by genetic complementation of the growth of the *Escherichia coli* *uxaB* knockout mutant strain on

*For correspondence. rodionov@burnham.org; Tel. (+1) 858 646 3100; Fax (+1) 858 795 5249.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. HHpred alignment of UxaE from *T. maritima* (TM0440) and the highest scoring protein with experimentally determined structure (tagatose-bisphosphate aldolase AgaY, pdb entry 1gvf). Experimental and predicted secondary structure is shown above and below the sequences: H – helix; E – strand; C, T – coil and turn; small letters denote less confident prediction.

Fig. S2. A. Phylogenetic tree of the UxaE family proteins. The tree was constructed using the maximum likelihood method (phym) from the PHYLIP package using the default parameters and visualized with Dendroscope. The numbers represent bootstrap values in percentages obtained from 1000 replicates.

B. Multiple alignment of UxaE family proteins analysed in this work. Conserved residues likely involved in the catalysis and discussed in the text are indicated.

Fig. S3. GC-MS analysis of hexuronic acid isomerization by *tm*_UxaC. The assay mixture containing either D-GlcA (A) or D-GalA (B) (black line) was incubated overnight with the recombinant *tm*_UxaC enzyme and showed the appearance of new peaks that correspond to either D-FruA (A) or D-TagA (B) respectively (red line). Retention times of hexuronic acid-specific peaks as determined on total ion count chromatogram: D-GlcA, Rt = 10.44 and 10.57 min; D-FruA, Rt = 10.15 and 10.27 min; D-GalA, Rt = 10.52 and 10.68 min; D-TagA, Rt = 10.39 and 10.51 min; latter peak obscured by D-GalA.

Fig. S4. Genetic complementation of *E. coli* *uxaB* and *uxuB* mutant strains by the pMH2T7-derived plasmids harbouring TM0442 (encoding *tm*_GntE), TM0440 (encoding *tm*_UxaE) or TM0808 gene (used as a negative control). The recombinant strains were spotted on M9 agar plates supplemented with 4 g l⁻¹ either D-GalA (A), D-GlcA (B) or D-gluconate (C) (used as a positive control) and incubated at 37°C for 24 h.

Fig. S5. Electron ionization mass spectra obtained for D-GlcA (A), D-GalA (B), D-FruA (C) and D-TagA (D). D-GalA and D-GlcA epimers have five characteristic m/z fragments (423, 393, 364, 333 and 305). D-TagA and D-FruA epimers are characterized by two abundant m/z fragments (364 and 277).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

galacturonate. An inferred novel galacturonate to mannonate catabolic pathway in *T. maritima* was reconstituted *in vitro* using a mixture of recombinant purified enzymes UxaE, UxaC and UxuB. Members of the newly identified UxaE family were identified in ~ 50 phylogenetically diverse heterotrophic bacteria from aquatic and soil environments. The genomic context of respective genes and reconstruction of associated pathways suggest that UxaE enzymatic and biological function remains conserved in all of these species.

Introduction

Carbohydrates are a primary source of carbon and energy for many microorganisms. Large variations in carbohydrate pools in different ecosystems are matched by variations in carbohydrate utilization pathways observed even for closely related microbial species. This variability includes alternative biochemical routes, non-orthologous gene replacements, functionally heterogeneous families of paralogues and other problematic issues that cannot yet be resolved by automated genome annotation pipelines. Additionally, a large fraction of microbial genes involved in these (as well as many other) pathways have no functional annotation (Brown and Babbitt, 2012).

Comparative genomic techniques based on genome context analysis often provide clues about the function of a gene even in cases when sequence similarity-based methods fail. Among these techniques, the analysis of chromosomal gene clustering, protein fusion events and phylogenomic co-occurrence profiles allow the inference of functional coupling between genes (Osterman and Overbeek, 2003). In addition, the comparative analysis of metabolic regulons is an efficient method for functional gene annotation as it can reveal sets of genes that are co-regulated by a common transcription factor (Rodionov, 2007). Functional genomics methods, such as gene expression microarrays, can also provide hints about function. We combined these techniques in a subsystems-based comparative genomic approach for reconstruction of metabolic pathways, identification of candidate (missing) genes for associated enzymes and transporters and experimental testing of selected bioinformatic predictions for poorly studied microorganisms (Yang *et al.*, 2006; 2008). This approach is particularly efficient for elucidation of novel sugar catabolic pathways that are often comprised of colocalized and co-regulated genes, as demonstrated recently by our systematic reconstruction of sugar utilization machinery in 19 species of *Shewanella* (Rodionov *et al.*, 2010). This effort led to reconstruction of 16 carbohydrate catabolic pathways, including 10 novel pathway variants, and determined novel functional roles for 62 previously unassigned genes.

An expansion of this approach to a deep-branched group of Thermotogales exposed even larger diversity of sugar catabolic pathways and associated genes. *Thermotoga* spp. are anaerobic fermentative bacteria that are capable of growing on various simple and complex carbohydrates including glucose, starch, cellulose, xylan and pectin (Connors *et al.*, 2006). Thus, a model organism in this genus, *Thermotoga maritima* MSB8, has a large proportion of genes (10–15%) involved in carbohydrate metabolism (Nelson *et al.*, 1999). However, exact functions of many of these genes remained unclear (Zhang *et al.*, 2009). This incomplete knowledge translated into some gaps in the first published model of the *T.*

maritima metabolic network (Zhang *et al.*, 2009). Improvement of this metabolic model is important for prospective bioenergy applications.

Pectins are complex polysaccharides that are present in the primary cell wall of plants (Ridley *et al.*, 2001). Many microorganisms are equipped with a set of pectinolytic enzymes including esterases, pectate lyases and polygalacturonases. D-galacturonate (D-GalA), the main monomeric constituent of pectin, serves as an important carbon source for pectin-degrading microorganisms. Several variants of D-GalA utilization pathways have been described (Richard and Hilditch, 2009). *Escherichia coli* uses the isomerase pathway for D-GalA utilization (Peekhaus and Conway, 1998), which produces 2-keto-3-deoxy-D-gluconate (KDG) via D-altronate (D-AltA) using three consecutive reactions: hexuronate isomerase UxaC, D-AltA oxidoreductase UxaB and D-AltA hydrolase UxaA. KDG is finally utilized by the sequential action of KDG kinase KdgK and aldolase KdgA (Fig. 1). The UxaC isomerase in *E. coli* is catalytically active on both hexuronic acid isomers, D-GalA and D-glucuronate (D-GlcA), and converts them to D-tagaturonate (D-TagA) and D-fructuronate (D-FruA) respectively (Ashwell *et al.*, 1960). In the D-GlcA catabolic pathway in *E. coli*, D-FruA is further converted to KDG via D-mannonate (D-ManA) by using D-ManA dehydrogenase UxuB and D-ManA dehydratase UxuA (Hickman and Ashwell, 1960; Smiley and Ashwell, 1960). The oxidative pathway of D-GalA utilization in *Agrobacterium tumefaciens* and *Delftia acidovorans* uses uronate dehydrogenase converting D-GalA to D-galactarate that is further catabolized to 2-oxoglutarate, whereas the non-phosphorylating D-GalA catabolic pathway in filamentous fungi proceeds through L-galactonate.

Two pectinolytic enzymes, the extracellular pectate lyase PelA (TM0433) and cytoplasmic exopolygalacturonidase PelB (TM0437), have been characterized in the hyperthermophilic bacterium *T. maritima* (Kluskens *et al.*, 2003; 2005). Although homologues of UxaC (TM0064), KdgK (TM0067) and KdgA (TM0066) were found in *T. maritima*, the D-GalA utilization pathway is incomplete because of the absence of UxaA and UxaB (Fig. 1). Here, we report identification and experimental characterization of a novel enzyme family, named UxaE, that catalyses the epimerization of D-TagA to D-FruA, a reaction that was not previously described in any organism. The elucidated novel variant of the D-GalA catabolic pathway provided a rationale for the known ability of *T. maritima* to grow on pectin as a sole carbon source (Kluskens *et al.*, 2003). By analysing genomic context of UxaE homologues in diverse bacterial genomes, we propose that most members of the UxaE family preserve their enzymatic function while contributing to several distinct variants of hexuronate catabolism.

Results

Bioinformatic analysis of hexuronate catabolic genes

Two gene clusters that are potentially involved in the pectin and hexuronate catabolism were identified in *T. maritima* and related species (Fig. 2). Using protein similarity searches, genome context analysis and previous experimental references, we collected functional roles of previously characterized proteins and inferred putative functions for other potential components of the hexuronate pathway (Table 1).

The first gene cluster *TM0430–443* contains 14 genes. Extracellular pectate lyase *PelA* degrades polygalacturonate to form unsaturated tri- and digalacturonates (Kluskens *et al.*, 2003). ABC-type transporter *AguEFG* binds digalacturonate *in vitro*, and thus can be involved in the uptake of pectin-derived oligogalacturonates (Nanavati *et al.*, 2006). Cytoplasmic enzyme *PelB* degrades saturated oligogalacturonates (Kluskens *et al.*, 2005). *Agu4A* has α -glucuronidase activity (Suresh *et al.*, 2002). *Gnd* catalyses the conversion of 6-phosphogluconate to ribulose-5-phosphate and CO₂ (Wang and Zhang, 2009). Functional roles of other genes in the *TM0430–443* gene cluster have not been studied before this work; thus, we used bioinformatic approaches to propose their potential functions, and targeted experiments to validate some of these novel assignments. *TM0439* (named *UxaR*) is a putative GntR family transcriptional regulator of hexuronic acid catabolic genes. *TM0441* (named *UxaD*) encodes an uncharacterized enzyme, which is 40% identical to the NADP-dependent 5-keto-D-gluconate reductase *IdnO* from *E. coli*. *TM0443* encodes a hypothetical FGGY-like sugar kinase that is 37% identical to the D-gluconate kinase *GntK* from *Bacillus subtilis*. *TM0442* and *TM0440* lack any characterized homologues. *TM0442* has a domain of unknown function (PF09861), whereas *TM0440* has no conserved domains according to the Pfam database. Functional roles of *TM0440* (named *UxaE*) and *UxaD* in the D-GalA catabolism were investigated in this study and are reported below.

The second gene cluster *TM0064–0069* contains six genes, of which only the aldolase gene *kdgA* was previously characterized (Griffiths *et al.*, 2002). *TM0064* is 35% identical to the hexuronate isomerase *UxaC* from *E. coli* (Ashwell *et al.*, 1960). The *kdgK*, *uxuB* and *uxuA* genes encode uncharacterized enzymes that are 27–40% identical to the respective D-GlcA catabolic enzymes from *E. coli*. The *kdgR* gene encodes a putative transcriptional regulator from the IclR family that is 34% identical to the KDG-responsive repressor *KdgR* from *E. coli* (Rodionov *et al.*, 2004). Using the comparative genomics approach implemented in the RegPredict web server (Novichkov *et al.*, 2010a), we identified a conserved 17 nt DNA motif located in the upstream regions of the *aguEFG*, *pelA*, *TM0436–443* and *kdgAK–uxuBA* genes in *T. maritima* and related species (Fig. 2). Based on the observation of shared regulatory sites, we propose that the hexuronic acid catabolic genes in *Thermotoga* spp. are co-regulated by the novel GntR-like transcription factor *UxaR*. Specific binding of the *T. maritima* *UxaR* regulator to the predicted operator sites was confirmed by *in vitro* DNA–protein binding assays (D.A. Rodionov, unpublished).

We used the HHpred server to search for distant homologues of *T. maritima* *UxaE* among proteins with experimentally determined spatial structures (pdb70_17Mar12 database). The first hit, with probability above 99%, was to the tagatose-bisphosphate aldolase *AgaY* from *E. coli* (pdb entry 1gvf). According to structural classification of proteins in the SCOP database, *AgaY* belongs to the aldolase superfamily in the TIM β/α barrel fold. Moreover, all hits with probability above 95% are class II aldolases (e.g. fructose-bisphosphate aldolase), and all hits above 70% probability are TIM barrels (e.g. arginine decarboxylase). This consistency of the results combined with a very high score of the first hit and conservation of metal ligands in the sequence alignment (Fig. S1) strongly suggests that *UxaE* is homologous to class II aldolases and adopts a TIM barrel structure.

We analysed the phylogenetic distribution of *uxaE* orthologues in microbial genomes using BLAST searches with TM0440 against a non-redundant protein sequences (nr) database. A single copy of the *uxaE* gene was detected in 35 bacterial genomes including eight Thermotogae species and in several other bacterial phyla including Acidobacteria, Bacteroidetes, Dictyoglomi, Firmicutes and Spirochaetes. We also identified a single case of the presence of *uxaE* in Archaea, in a genome of the halophilic archaeon *Halorhabdus utahensis*. Phylogenetic analysis of the identified UxaE proteins suggests the involvement of multiple horizontal gene transfer events in the evolution of the UxaE family. For instance, *uxaE* genes in various species from the Spirochaetes and Firmicutes phyla are scattered along the phylogenetic tree (Fig. S2).

The genomic context of identified *uxaE* genes was analysed to reveal the putative metabolic pathways and functional partners of UxaE-like enzymes. In most cases (31 out of 36 genomes), *uxaE* is colocalized on the chromosome with either known (*uxaC*, *uxuAB*, *uxaAB*, *kdgKA*, *gntK*, *gnd*) or predicted (*uxaD*, *gntE*) genes from the hexuronic acid or D-gluconate utilization pathways (Fig. 3). The genomic neighbourhood analysis revealed the *uxaD*, *gntK*, *gnd* and *uxaC* genes as the most common genomic partners of *uxaE*. These genes were found within the *uxaE*-containing chromosomal clusters in 19, 14, 12 and 11 genomes respectively. The phylogenetic co-occurrence profiles identified UxaC, KdgA, KdgK, UxaD, GntK and Gnd as the most frequent partners of UxaE in the analysed genomes, and were found in 34, 32, 29, 27, 26 and 26 cases respectively. Interestingly, UxuB and UxaD have a partially complementary pattern of occurrence in these genomes, so that any UxaE-containing genome has either one of these proteins, or both of them.

In summary, the presence of *uxaC*, *kdgK* and *kdgA* homologues, and their genomic colocalization and potential co-regulation with the pectin utilization genes *pelA*, *pelB* and *aguEFG* suggest the existence of a functional D-GalA catabolic pathway in *T. maritima*. This is consistent with the ability of *T. maritima* to grow on pectin, a D-GalA-containing polysaccharide, as a sole carbon and energy source (Kluskens *et al.*, 2003). The extracellular pectate lyase PelA degrades polygalacturonate to form unsaturated trigalacturonate, which is further converted by the same enzyme to unsaturated digalacturonate (Kluskens *et al.*, 2003). The ABC-type transporter AguEFG was proposed to be involved in the uptake of oligogalacturonates derived from pectin degradation. The extracellular substrate-binding protein AguE binds digalacturonate *in vitro* (Nanavati *et al.*, 2006). The cytoplasmic exopolygalacturonidase PelB further degrades saturated oligogalacturonates by removing monogalacturonate units from their non-reducing end (Nanavati *et al.*, 2005). It is not clear how unsaturated oligogalacturonates are converted to their saturated derivatives in *T. maritima*.

However, the cytoplasmic gal catabolic pathway of *T. maritima* was not resolved before this work. The *Thermotoga* genomes have no gene candidates for the UxaA- and UxaB-catalysed reactions that link D-TagA to KDG in the D-GalA catabolic pathway in *E. coli* (Fig. 1). These observations allowed us to propose that *Thermotoga* uses a novel variant of the D-GalA catabolic pathway. Below, we describe the experimental results that allowed us to elucidate this novel pathway, which employs a previously unknown enzymatic activity D-TagA/D-FruA epimerase performed by a novel UxaE protein family.

Experimental characterization of hexuronate catabolic enzymes

To characterize the novel hexuronic acid catabolic enzymes in *T. maritima*, the recombinant proteins *tm_UxaC* (TM0064), *tm_UxaE* (TM0440), *tm_UxuB* (TM0068) and *tm_UxaD* (TM0441) were produced and characterized *in vitro* by direct and coupled enzymatic assays using spectrophotometry and GC-MS.

Hexuronate isomerase—In *E. coli*, the first step of D-GalA catabolism is catalysed by the hexuronic acid isomerase UxaC. The *ec_UxaC* enzyme has comparable specificities towards D-GalA and D-GlcA (K_m 3.7 mM and 1.6 mM respectively) (Ashwell *et al.*, 1960), and thus operates as a hub contributing to catabolic pathways of both hexuronic acids. Metabolic reconstruction of the hexuronate catabolic network in *T. maritima* (Fig. 1) suggested a similar dual specificity for *tm_UxaC* (TM0064). This conjecture was verified using coupled colorimetric assays with two downstream NADH-dependent dehydrogenases: (i) altronate oxidoreductase *ec_UxaB* reducing D-TagA to D-AltA and (ii) mannonate dehydrogenase *tm_UxuB* (TM0068) reducing D-FruA to D-ManA. The depletion of NADH was measured in both coupled assays, thus confirming *tm_UxaC* activity, while no activity was detected in the absence of *tm_UxaC*. Kinetic parameters obtained for the recombinant *tm_UxaC* revealed comparable affinity for both substrates but 7.8-fold higher catalytic efficiency for D-GalA versus D-GlcA (Table 2). These observations are consistent with the existence of the functional catabolic pathway for D-GalA in *T. maritima*, starting from its conversion to D-TagA. Therefore, the next step in this novel pathway variant (lacking UxaB enzyme) was expected to employ a yet-unknown enzyme acting on D-TagA.

Mannonate dehydrogenase—Although the *ec_UxuB* enzyme is apparently committed to the D-GlcA utilization pathway, it was reported to show a comparable activity on both D-FruA and D-TagA (Hickman and Ashwell, 1960). Therefore, its orthologue in *T. maritima* (*tm_UxuB* or TM0068) could be considered a candidate enzyme for the next step of D-GalA catabolism (in addition to its obvious role in the conventional D-GlcA pathway, which is present in *T. maritima*, see Fig. 1). However, a stringent substrate preference for D-FruA displayed by *tm_UxuB* argues strongly against this possibility. Indeed, the recombinant purified *tm_UxuB* protein showed high catalytic efficiency on D-FruA substrate but no activity on D-TagA, D-GalA or D-GlcA (data not shown). In the reverse direction, *tm_UxuB* also demonstrated high specificity towards D-ManA (Table 2, *tm*Fig. 4A). Other features of *tm_UxuB* that we established are its strict cofactor preference, NAD(H) but not NADP(H), and general thermostability (in a broad range of temperatures 30–80°C).

Fructuronate reductase—The *tm_UxaD* (TM0441) protein is distantly homologous to NADP-dependent 5-keto-D-gluconate reductase GNO from *Gluconobacter oxydans* (Klasen *et al.*, 1995). Based on genomic context it could be considered a candidate enzyme for the D-GalA utilization pathway. However, the purified recombinant *tm_UxaD* protein showed only a weak activity on D-FruA (Fig. 4B) and no activity on D-TagA, D-GalA or D-GlcA (data not shown). Similarly to *G. oxydans* GNO, D-FruA reductase activity of *tm_UxaD* was strictly NADPH (and not NADH)-dependent. Almost no activity could be detected in the reverse direction (from D-ManA). This is in contrast with reversible NAD(H)-dependent interconversion of D-FruA and D-ManA by a canonical UxuB enzyme from both *E. coli* and

T. maritima. The ability of *tm*_UxaD to reduce D-FruA to D-ManA initially postulated from the colorimetric detection of NADPH to NADP conversion was directly confirmed by GC-MS analysis (Fig. 5).

The physiological significance (if any) of this apparent functional redundancy – the existence of the two enzymes (UxuB and UxaD) catalysing the same reaction (albeit with distinct cofactors) – is unclear. Further analysis of *tm*_UxaD revealed activity on 5-keto-D-gluconate, which is characterized by kinetic parameters comparable to those on D-FruA as substrate (Table 2). However, the physiological relevance of 5-keto-D-gluconate as a substrate for UxaD in *T. maritima* is questionable.

Tagaturonate/fructuronate epimerase—To assess the possible enzymatic activity of TM0440, we used GC-MS profiling of reaction mixtures obtained by an overnight incubation of commercially available hexuronates (and related metabolites) with the purified recombinant TM0440 protein taken alone or in combination with other proteins from this catabolic network. While incubation of D-GalA or D-GlcA with TM0440 alone did not produce any new peaks on the chromatogram, the addition to the mixture of *tm*_UxaC enzyme led to notable changes. Thus, incubation of D-GlcA with TM0440 and *tm*_UxaC led to appearance of novel peaks, detected as D-FruA, D-TagA and D-GalA (data not shown). For further analysis we used D-FruA, which was obtained enzymatically from D-GlcA (see *Experimental procedures*). After incubation of TM0440 with the D-FruA/D-GlcA mixture overnight, we observed the decrease of two peaks corresponding to D-FruA and the appearance of two new peaks with retention times 10.39 and 10.51 min (Fig. 6). The first peak at 10.39 min can be confidently assigned as D-TagA based on the GC-MS analysis of *tm*_UxaC-driven isomerization of D-GalA to D-TagA (Fig. S3). The second peak at 10.51 min is attributed to D-TagA epimer by analysis of electron ionization mass spectra (Fig. S4). These results (later supported by focused coupled assays and complementation experiments, see below) provided the first evidence that TM0440 can catalyse an interconversion of D-FruA and D-TagA.

To our knowledge, the respective enzymatic activity, which may be termed tagaturonate–fructuronate epimerase, has not been previously detected or even postulated in any species. Notably, this activity provides a missing link in this novel variant of D-GalA catabolism merging it with the canonical D-GlcA pathway at the stage of D-FruA (Fig. 1). Based on the naming convention for D-GalA catabolism, we propose the common name UxaE for this novel enzyme family (represented by TM0440 in *T. maritima*).

For further enzymatic characterization of *tm*_UxaE we developed a colorimetric assay coupling the conversion of D-FruA to D-TagA with NADH-dependent reduction of D-TagA to D-AltA by altronate dehydrogenase from *E. coli* (*ec*_UxaB) (Fig. 7A). First, we tested enzymatic activity of the purified recombinant *ec*_UxaB coupling enzyme on hexuronic acids (D-GalA, D-GlcA, D-FruA and D-TagA) and confirmed its previously reported stringent specific activity towards D-TagA (Hickman and Ashwell, 1960). This assay allowed us to confirm the robust enzymatic activity of *tm*_UxaE and establish its steady-state kinetic parameters (Table 1). The reaction in the opposite direction, from D-TagA to D-FruA, which is likely of higher physiological importance for the *T. maritima* hexuronate

network, was confirmed using a similar coupled assay with *tm_UxuB* (Fig. 7B). Due to thermostability of *tm_UxuB*, this assay was also used to assess the temperature dependence of UxaE activity. The peak of UxaE activity is observed at 60°C, whereas the enzyme still showed ~40% of its maximal activity at 90°C (Fig. 7B).

Genetic complementation in *E. coli*

A functional role of *tm_UxaE* from *T. maritima* was additionally confirmed by genetic complementation in the *E. coli* *uxaB* and *uxuB* mutants. In contrast to the wild type, the *uxaB* strain of *E. coli* does not grow on minimal medium with D-GalA as a sole carbon and energy source. Transformation by the TM0440 expression construct encoding *tm_UxaE* restored this ability (Fig. S4). Similarly, the *uxuB* strain of *E. coli* does not grow on D-GlcA in minimal medium, while the introduction of the TM0440 gene into this strain restores this ability (Fig. S4). We conclude that the D-TagA/D-FruA epimerase activity of UxaE can compensate either for the lack of D-AltA dehydrogenase activity when grown on D-GalA or for the lack of D-ManA dehydrogenase activity when grown on D-GlcA, serving as a shunt linking two hexuronic acid catabolic pathways.

In vitro pathway reconstitution in *T. maritima*

We further validated the novel variant of the D-GalA catabolic pathway proposed for *T. maritima* by *in vitro* reconstitution of the three-step biochemical conversion of D-GalA to D-ManA. The reaction mixtures contained 2 mM D-GalA, 0.2 mM NADH, and combinations of the purified recombinant enzymes *tm_UxaC* (0.1 µg), *tm_UxaE* (1 µg) and *tm_UxuB* (0.5 µg). The coupled NADH-detecting assay was conducted at 45°C. The decrease in NADH absorbance was detected only in the presence of all three recombinant enzymes, while no NADH depletion was measured with any two of these enzymes.

Discussion

The comparative genomics analysis and metabolic reconstruction revealed a novel variant of the hexuronic acid catabolic pathway in hyperthermophilic bacteria from the deep-branched genus of *Thermotoga*. A previously uncharacterized protein family (as represented by TM0440 in *T. maritima*) was implicated as a gene candidate for the novel D-GalA pathway variant in many diverse bacteria. We combined bioinformatics and experimental techniques to assess enzymatic function and functional role of TM0440, which was found to catalyse the epimerization of D-TagA to D-FruA, a novel reaction in the catabolism of hexuronic acids (Fig. 1). This enzymatic activity, which has not been previously described in any species, provides a shunt between D-GalA and D-GlcA utilization pathways, closing the gap in the reconstruction of the hexuronate catabolic network of *T. maritima*. Orthologues of the novel D-TagA epimerase, termed UxaE, are present in all analysed *Thermotoga* spp. (Fig. 2), and in several other phyla of Bacteria and Archaea (Fig. 3). In public databases *uxaE*-like genes are annotated as putative and no homology with any other functionally assigned genes could be found by BLAST. Distant homology analysis using sequence profile methods predicted UxaE as a distant homologue of class II aldolases that adopt the TIM barrel fold. The inferred structural similarity along with the conservation of the key active site residues provided a foundation for prediction of the UxaE catalytic mechanism.

Possible catalytic mechanism for *UxaE* epimerase

Class II aldolases utilize a metal cofactor, mostly Zn^{2+} , to stabilize the enolate intermediate during catalysis. In the active site of the AgaY aldolase, a distant homologue of UxaE, Zn^{2+} is chelated by three histidines, His⁸³, His¹⁸⁰ and His²⁰⁸ (Hall *et al.*, 2002). Two of these metal chelating residues are conserved in all UxaE proteins (corresponding to His¹⁶² and His³⁴¹ in TM0440), whereas His¹⁸⁰ in AgaY is substituted by a conserved Lys³⁰⁸ residue in UxaE (Fig. S2). This suggests that UxaE epimerase also utilizes a divalent metal ion for catalysis. The catalytically important Asp⁸² in AgaY binds to the C3 hydroxyl hydrogen of tagatose-bisphosphate. This residue corresponding to Asp¹⁶¹ in TM0440 is also conserved in all UxaE proteins (Figs S1 and S2), possibly providing similar interactions with the C3 hydroxyl hydrogen of D-TagA/D-FruA (Fig. 8). Comparison with another class II aldolase, the fructose-bisphosphate aldolase Fba from *Thermus aquaticus*, reveals Glu¹³² residue that interacts with the C2 keto-group of fructose-bisphosphate (Izard and Sygusch, 2004). The Glu¹³² residue in Fba corresponds to Glu²⁶⁶ in TM0440, and this position is absolutely conserved in all UxaE proteins (Fig. S2). Thus, we hypothesized that Asp¹⁶¹ and Glu²⁶⁶ residues in UxaE are possibly involved in its catalytic mechanism.

This mechanism is expected to share common features with class II aldolases (such as AgaY and Fba) rather than with other (structurally distinct and relatively well studied) families of carbohydrate epimerases. By analogy, we hypothesize that UxaE-catalysed D-TagA/D-FruA epimerization may proceed via carbon-carbon bond cleavage, similar to the class II aldolase mechanism. The epimerization via carbon-carbon bond cleavage was previously described for L-ribulose-5-phosphate epimerase (AraD), which is functionally and structurally related to class II aldolases (Luo *et al.*, 2001). The retroaldol/aldol mechanism of the AraD-catalysed epimerization requires the presence in the active site of two residues providing a general acid and a general base for catalysis, which is expected to be true for UxaE. The proposed catalytic mechanism of D-TagA epimerase is schematically shown in Fig. 8. Asp¹⁶¹ likely interacts with the C3 hydroxyl group of D-TagA and deprotonates it. A divalent cation (likely Zn^{2+}) chelated by His¹⁶² and His³⁴¹ (the third Zn^{2+} ligand, His¹⁸⁰ of AgaY is replaced by Lys³⁰⁸) possibly interacts with the C5 keto and C4 hydroxyl groups of D-TagA, thereby contributing to the cleavage of the adjacent C3-C4 bond. A carboxyl group of Glu²⁶⁶ then may promote the rotation and re-protonation of the C3 aldehyde group in the intermediate and the restoration of the broken C3-C4 bond to form D-FruA. The involvement of the two carboxylic acid residues in the catalytic mechanism of UxaE may occur in the opposite order, when Glu²⁶⁶ initiates the reaction by deprotonation of D-TagA, and Asp¹⁶¹ completes the reaction by promoting the formation of D-FruA.

Novel variants of hexuronic acid catabolic pathways in bacteria

The analysis of the hexuronic acid utilization pathway in *T. maritima* and related species has allowed us to significantly improve the quality of genomic annotations and to accurately infer associated metabolic and regulatory networks. The novel tagaturonate epimerase UxaE is highly conserved in at least 35 bacterial genomes representing diverse taxonomic groups. UxaE catalyses a novel enzymatic activity, which provides an upstream link between the D-GalA and D-GlcA utilization routes. Thus, the utilization of both hexuronic acids in *T. maritima* proceeds through D-FruA and D-ManA intermediates (Fig. 1). Identification of a

novel enzymatic shunt in hexuronic acid catabolism suggests a possible novel pathway variants in bacterial genomes containing UxaE orthologues.

Genomics-based metabolic reconstruction of hexuronate catabolism in such genomes revealed a diversity of associated pathway variants (Fig. 9). In all of these variants a bifunctional D-GalA/D-GlcA isomerase UxaC catalyses the first step of formation of D-TagA and/or D-FruA intermediates that can be interconverted via UxaE-catalysed epimerization. The first, major, pathway variant (variant A) is characterized by the presence of signature enzymes UxuB (or UxaD) and UxuA. This variant, which includes the flux through D-ManA and KDG, was identified in 17 out of 27 analysed genomes. The second pathway variant (variant B) comprised of signature enzymes UxaB and UxaA proceeds via D-AltA and KDG. It was identified in a subset of six bacteria with a complete UxuBA pathway. For instance, in *Halanaerobium hydrogeniformans* the Halasa_0901–Halsa_0913 gene cluster encodes all enzymes required for conversion of D-GalA and D-GluA to KDG via both pathway variants.

The genome context analysis of other UxaE-encoding genomes revealed the existence of another potential route for D-ManA utilization via the pentose phosphate pathway in 19 bacteria including *T. maritima* and other *Thermotoga* spp. (Fig. 9). This novel pathway variant (variant C) has two signature genes, an uncharacterized gene of unknown function (*TM0442*, termed *gntE*) and D-gluconate kinase gene (*TM0443*, termed *gntK*), that are colocalized with other D-GalA/D-GlcA catabolic genes in a large number of analysed genomes (Fig. 3). *In vitro* characterization of the purified *tm_GntK* enzyme from *T. maritima* confirmed its activity on D-gluconate (Rodionova *et al.*, 2012). In *Thermotoga* and many other bacteria, the *gntE*–*gntK* genes are located in a cluster with the *gnd* gene encoding the first enzyme of the pentose phosphate pathway. Many of the above species, e.g. *Dictyoglomus*, are lacking both the UxuA- and the UxaA-mediated routes for hexuronic acid catabolism. Based on this genomic evidence and the absence of candidate gluconate transporters we suggest the existence of yet another metabolic route for hexuronic acid utilization that proceeds through D-gluconate and enters the pentose phosphate pathway (Fig. 9). In this novel pathway, the interconversion between D-ManA and D-gluconate can be mediated by a yet uncharacterized enzyme encoded by the hypothetical gene *gntE*. We are currently working on *in vitro* characterization of *tm_GntE* from *T. maritima*. Interestingly, this novel variant of the D-GalA/D-GlcA catabolic pathway via D-gluconate was identified as the only functional pathway variant in 10 genomes possessing the *uxaE* epimerase gene.

In summary, by combining a subsystems-based reconstruction of metabolic and regulatory network of hexuronate catabolism with the focused experimental characterization of identified gene candidates, we identified a novel enzyme, D-TagA/D-FruA epimerase. This enzyme, termed UxaE, catalyses a previously unknown biochemical transformation, a shunt bridging D-GalA and D-GlcA utilization pathways. A comparative metabolic reconstruction in UxaE-containing bacterial genomes revealed a substantial diversity of associated pathway variants expanding our knowledge of hexuronate catabolic networks.

Experimental procedures

Genome resources and bioinformatics tools

Identification of orthologues in closely related genomes and gene neighbourhood analysis were performed in MicrobesOnline (<http://www.microbesonline.org>) (Dehal *et al.*, 2010). The comparative genomic analysis of the D-GalA and D-GluA utilization subsystem was performed using the SEED genomic platform (<http://pubseed.theseed.org>), which supports the encoding and projection of a metabolic subsystem across an entire collection of integrated genomes and allows identification of non-orthologous gene displacements in the analysed metabolic subsystems (Overbeek *et al.*, 2005). Identification of candidate regulatory sites in genomic sequences and regulon reconstruction were performed using the RegPredict software (Novichkov *et al.*, 2010a). The details of the reconstructed UxaR regulon are captured and displayed in the RegPrecise database (<http://regprecise.lbl.gov/>) (Novichkov *et al.*, 2010b). The phylogenetic tree of UxaE proteins was built using the maximum likelihood method (phylml) from the PHYLIP package (Felsenstein, 1996) using the default parameters and the bootstrap values from 1000 replicates. Cladogram and phylogram of the UxaE tree were visualized with Dendroscope (Huson *et al.*, 2007). Orthologues of UxaE and other hexuronate catabolic genes were identified by NCBI BLAST searches in the 'nr' database (Johnson *et al.*, 2008) and confirmed by phylogenetic tree construction using the FastTree algorithm in MicrobesOnline (Price *et al.*, 2009). The long range homology analysis and fold prediction were performed using the HHpred server (Soding *et al.*, 2005), Pfam database of protein families (Punta *et al.*, 2012) and SCOP database of protein structures (Murzin *et al.*, 1995).

Bacterial strains, plasmids and chemicals

Escherichia coli strain DH5 α (Invitrogen, Carlsbad, CA, USA) was used for gene cloning, and BL21 and BL21 (DE3) (Gibco-BRL, Rockville, MD, USA) were used for protein over-expression. The *E. coli* BL21 strains containing pMH2T7-derived plasmids that harbour the *T. maritima* TM0440 (*tm_uxaE*), TM0068 (*tm_uxuB*), TM0441 (*tm_uxaD*) or TM0064 (*tm_uxaC*) genes with an arabinose promoter and His₆ tag were a kind gift from Dr S. Lesley at the Joint Center for Structural Genomics (La Jolla, CA, USA). The *E. coli uxaB* gene (*ec_uxaB*) with the isopropyl 1-thio- β -D-galactopyranoside (IPTG)-inducible promoter and the His₆ tag was from the ASKA gene expression library (Kitagawa *et al.*, 2005). The pSMT3 expression vector (Mossessova and Lima, 2000) used for the expression of TM0440 was a kind gift of Dr Lima from Cornell University. For TM0440 gene cloning, two specific primers containing the BamHI and HindIII restriction sites were designed, 5'-ccatcAT GGATCCATGGTCTTGAAAGTGTT, and 5'-aattaaaagc ttCACCCTCCAGCAGAT. The resulting plasmid contained the *T. maritima uxaE* gene with His₆ and Smt3 protein tags. For genetic complementation in *E. coli*, we used the knockout mutant *uxaB* strain from the KEIO collection (a kind gift from Dr Mori) (Baba *et al.*, 2006). Enzymes for PCR and DNA manipulations were from New England Biolabs (Beverly, MA, USA). Plasmid purification kits were from Promega (Madison, WI, USA). PCR purification kits and nickel-nitrilotriacetic acid (Ni-NTA) resin were from QIAGEN (Valencia, CA, USA). Oligonucleotides for PCR and sequencing were synthesized by Sigma-Genosys (Woodlands, TX, USA). D-GlcA, NADH and NADPH were purchased from Sigma-Aldrich (St Louis,

MO, USA). D-GalA and D-mannonate-1,4-lactone were from Carbosynth (Berkshire, UK). D-ManA was produced from D-mannonate-1,4-lactone by alkaline hydrolysis. D-FruA and D-TagA are not commercially available intermediates of the hexuronic acid catabolic pathway. D-FruA and D-TagA were enzymatically prepared by *tm_UxaC* from D-GlcA and D-GalA, respectively, in the reaction mixture containing 200 mM Tris-HCl buffer (pH 7), 10 μ M ZnSO₄ and 30 mM of the corresponding hexuronic acid. The product concentration was determined by GC-MS (see below).

Protein purification

Recombinant proteins *tm_UxaC*, *tm_UxuB*, *tm_UxaD* and *ec_UxaB* containing an N-terminal His₆ tag were overexpressed in *E. coli* in 50 ml volume culture and purified using Ni²⁺-chelating chromatography, as previously described (Yang *et al.*, 2008). *tm_UxaE* was produced as a Smt3–His₆-tagged recombinant protein, from a larger-volume culture (1 l) using a standard two-step procedure, Ni-NTA chromatography with gradient elution followed by gel filtration in the AKTA-FPLC system (Gerdes *et al.*, 2006). After purification, N-terminal polyhistidine Smt3 fusion tag was removed from the recombinant UxaE protein by incubation with the Ulp-4 protease in 10 mM Hepes buffer (pH 7.2) containing 100 mM NaCl and 2 M urea for 2 days at room temperature. The UxaE monomer has a predicted molecular weight 41 kDa. Molecular mass of the purified recombinant UxaE enzyme was calculated by gel filtration to be ~ 78 kDa, which corresponds to the dimer.

Enzyme assays

Hexuronate isomerase activity of *tm_UxaC* on D-GlcA and D-GalA was analysed in a colorimetric assay, coupled with NADH-dependent dehydrogenases that are specific to D-FruA (*tm_UxuB*) or D-TagA (*ec_UxaB*) respectively. The oxidation of NADH was monitored at 37°C by decrease in the absorbance at 340 nm. The purified recombinant enzymes, *tm_UxaC* (0.1 μ g), *tm_UxuB* (4 μ g) or *ec_UxaB* (4 μ g), were added to 100 μ l of reaction mixture containing 100 mM Tris-HCl (pH 7), 10 μ M ZnSO₄ and 0.2 mM NADH. Assays were started by the addition of substrate (either D-GlcA or D-GalA respectively) in a concentration range of 1–8 mM.

Oxidoreductase activity of *tm_UxuB* on different sugars was tested by a direct NADH-detecting assay. The purified recombinant enzyme *tm_UxuB* (0.1 μ g) was incubated with increasing concentrations, from 0.3 to 5 mM, of a potential substrate, which was either D-FruA, D-TagA or 5-keto-D-gluconate, in 100 μ l of reaction mixture containing 10 μ M ZnSO₄, 0.25 mM NADH and 100 mM Tris-HCl (pH 8.5). The decrease in NADH was followed by reduction in absorbance at 340 nm. The activity of *tm_UxuB* in the reverse reaction on D-ManA was tested in the same assay mixture supplemented by 2 mM NAD⁺ instead of NADH.

Fructuronate reductase activity of *tm_UxaD* was assessed by a direct NADPH-detecting assay. The purified recombinant *tm_UxaD* enzyme (0.1 μ g) was incubated in 100 μ l of reaction mixture containing 10 mM CaCl₂, 0.25 mM NADPH and 100 mM Tris-HCl (pH 8.0) and supplemented with D-FruA at increasing concentrations from 1.5 to 17.5 mM. NADPH consumption was followed by change in absorbance at 340 nm.

Tagaturonate epimerase activity of *tm*_UxaE was characterized by two alternative coupled assays: (i) with *ec*_UxaB dehydrogenase for the conversion of D-FruA to D-TagA and (ii) with *tm*_UxuB dehydrogenase for the conversion of D-TagA to D-FruA (Fig. 1). Since none of these interconverting metabolites (D-TagA or D-FruA) are commercially available, they were generated *in situ* from their respective metabolic precursors using *tm*_UxaC enzyme (see next section). Each assay mixture was supplemented by 10 μ M ZnSO₄, 0.2 mM NADH and 100 mM Tris-HCl (pH 8.5). The consumption of NADH was followed at 340 nm. The first assay mixture (100 μ l) containing the recombinant proteins *tm*_UxaE (0.3 μ g) and *ec*_UxaB (4 μ g) was used to evaluate kinetic parameters of *tm*_UxaE over the range of D-FruA concentrations (0.6 to 3.2 mM) at room temperature. In the second assay (100 μ l of reaction mixture with 0.3 μ g of *tm*_UxaE and 4 μ g of *tm*_UxuB), a fixed saturating concentration of D-TagA (5 mM) was used to assess the temperature dependence of *tm*_UxaE in the range from 30° to 90°C.

To analyse the temperature dependence of the UxaE-catalysed reaction and to demonstrate its reversibility, we used a coupled colorimetric assay of UxaE with the thermostable enzyme *tm*_UxuB from *T. maritima*. The assay mixture containing 0.2 mM NADH and 2 mM D-TagA was pre-incubated during 20 min at different temperatures (30–90°C). Then the *tm*_UxaE and *tm*_UxuB proteins were added to the assay mixture and the depletion of NADH was followed by increase in absorbance at 340 nm at the same temperature in the spectrophotometer.

GC-MS analysis of hexuronic acids

Interconversions between hexuronic acid isomers were monitored by GC-MS. Samples from enzymatic assay mixtures were dried in a vacuum centrifuge at room temperature, and derivatized at 85°C for 20 min with 70 μ l of pyridine with 50 mg ml⁻¹ methoxylamine, and then at 80°C for 60 min with 60 μ l of *N,O*-Bis(trimethylsilyl)trifluoroacetamide with 1% trimethylchlorosilane from Thermo Fisher Scientific (Rockford, IL, USA). After derivatization, the samples were centrifuged for 1 min at 14 000 r.p.m. and the supernatant was transferred to vials for GC-MS analysis. A QP2010 Plus GC-MS instrument from Shimadzu (Columbia, MD, USA) was programmed with an injection temperature of 250°C, injection split ratio 1/10, with injection volume 1 μ l. Gas chromatography oven temperature started at 130°C for 4 min, rising to 250°C at 10°C min⁻¹ with a final hold at this temperature for 2 min. Flow rate of helium carrier gas was 50 cm s⁻¹. The column was a 15 m $\times$ 0.25 mm $\times$ 0.25 μ m SHRXI-5ms from Shimadzu. The GC-MS interface temperature was 300°C and ion source temperature was 200°C, with 70 V : 150 μ A ionization voltage : current. The mass spectrometer was set to scan *m/z* range 50–700, with $\sim$ 1 kV detector sensitivity. After derivatization, each hexuronic acid produced two peaks on the total ion count chromatogram that correspond to syn- and anti-methoximes (Fig. S3). Electron ionization spectra obtained for hexuronic acid isomers were used as standards in GC-MS analysis (Fig. S5). The electron ionization spectra for D-TagA and D-FruA epimers were similar to each other but different from the spectra obtained for D-GalA and D-GlcA.

Genetic complementation

The *uxaB* and *uxuB* mutant strains of *E. coli* K-12 BW25113 from the KEIO knockout collection (Baba *et al.*, 2006) were used for the complementation analysis. The phenotype of the *uxaB* mutant was confirmed by growth experiments in M9 minimal medium supplemented by 20 mM D-GalA used as a sole carbon source. The phenotype of the *uxuB* mutant was confirmed by growth experiments in M9 minimal medium supplemented with 20 mM D-GlcA used as a sole carbon source. The pMH2T7-derived plasmids harbouring TM0440 (encoding *tm_UxaE*) or TM0808 gene (used as a negative control) were introduced into *E. coli uxaB* and *uxuB* mutants. The obtained complementation strains were pre-cultivated on LB medium, then washed twice and diluted to the same optical density ($OD_{600} \sim 10^{-4}$) using the M9 minimal medium without any carbon source. The recombinant strains (5 μ l) were spotted on M9 agar plates supplemented with 4 g l⁻¹ either D-GalA, D-GlcA or D-gluconate (used as a positive control) and incubated at 37°C for 24 h.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was partially supported by the US Department of Energy, Office of Science (Biological and Environmental Research), as part of Genomic Science Program contracts DE-FG02-08ER64686 and DE-SC0004999 with Sanford-Burnham Medical Research Institute. In addition, the project was supported by COMBEX as a part of GO grant from National Institute of General Medical Sciences (1RC2GM092602-01) and by Russian Foundation for Basic Research (10-04-01768).

References

- Ashwell G, Wahba AJ, Hickman J. Uronic acid metabolism in bacteria. I. Purification and properties of uronic acid isomerase in *Escherichia coli*. *J Biol Chem*. 1960; 235:1559–1565. [PubMed: 13794771]
- Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*. 2006; 2:2006.0008.
- Brown SD, Babbitt PC. Inference of functional properties from large-scale analysis of enzyme superfamilies. *J Biol Chem*. 2012; 287:35–42. [PubMed: 22069325]
- Conners SB, Mongodin EF, Johnson MR, Montero CI, Nelson KE, Kelly RM. Microbial biochemistry, physiology, and biotechnology of hyperthermophilic *Thermotoga* species. *FEMS Microbiol Rev*. 2006; 30:872–905. [PubMed: 17064285]
- Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, Chivian D, et al. MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res*. 2010; 38:D396–D400. [PubMed: 19906701]
- Felsenstein J. Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol*. 1996; 266:418–427. [PubMed: 8743697]
- Gerdes SY, Kurnasov OV, Shatalin K, Polanuyer B, Sloutsky R, Vonstein V, et al. Comparative genomics of NAD biosynthesis in cyanobacteria. *J Bacteriol*. 2006; 188:3012–3023. [PubMed: 16585762]
- Griffiths JS, Wymer NJ, Njolito E, Niranjanakumari S, Fierke CA, Toone EJ. Cloning, isolation and characterization of the *Thermotoga maritima* KDPG aldolase. *Bioorg Med Chem*. 2002; 10:545–550. [PubMed: 11814840]

- Hall DR, Bond CS, Leonard GA, Watt CI, Berry A, Hunter WN. Structure of tagatose-1,6-bisphosphate aldolase. Insight into chiral discrimination, mechanism, and specificity of class II aldolases. *J Biol Chem.* 2002; 277:22018–22024. [PubMed: 11940603]
- Hickman J, Ashwell G. Uronic acid metabolism in bacteria. II. Purification and properties of D-altronic acid and D-mannonic acid dehydrogenases in *Escherichia coli*. *J Biol Chem.* 1960; 235:1566–1570. [PubMed: 14401695]
- Huson DH, Richter DC, Rausch C, DeZulian T, Franz M, Rupp R. Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinformatics.* 2007; 8:460. [PubMed: 18034891]
- Izard T, Sygusch J. Induced fit movements and metal cofactor selectivity of class II aldolases: structure of *Thermus aquaticus* fructose-1,6-bisphosphate aldolase. *J Biol Chem.* 2004; 279:11825–11833. [PubMed: 14699122]
- Johnson M, Zaretskaya I, Raytselis Y, Merezukh Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res.* 2008; 36:W5–W9. [PubMed: 18440982]
- Kitagawa M, Ara T, Arifuzzaman M, Ioka-Nakamichi T, Inamoto E, Toyonaga H, Mori H. Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res.* 2005; 12:291–299. [PubMed: 16769691]
- Klasen R, Bringer-Meyer S, Sahm H. Biochemical characterization and sequence analysis of the gluconate: NADP 5-oxidoreductase gene from *Gluconobacter oxydans*. *J Bacteriol.* 1995; 177:2637–2643. [PubMed: 7751271]
- Kluskens LD, van Alebeek GJ, Voragen AG, de Vos WM, van der Oost J. Molecular and biochemical characterization of the thermoactive family 1 pectate lyase from the hyperthermophilic bacterium *Thermotoga maritima*. *Biochem J.* 2003; 370:651–659. [PubMed: 12443532]
- Kluskens LD, van Alebeek GJ, Walther J, Voragen AG, de Vos WM, van der Oost J. Characterization and mode of action of an exopolysaccharuronase from the hyperthermophilic bacterium *Thermotoga maritima*. *FEBS J.* 2005; 272:5464–5473. [PubMed: 16262687]
- Luo Y, Samuel J, Mosimann SC, Lee JE, Tanner ME, Strynadka NC. The structure of L-ribulose-5-phosphate 4-epimerase: an aldolase-like platform for epimerization. *Biochemistry.* 2001; 40:14763–14771. [PubMed: 11732895]
- Mossesso E, Lima CD. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol Cell.* 2000; 5:865–876. [PubMed: 10882122]
- Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol.* 1995; 247:536–540. [PubMed: 7723011]
- Nanavati DM, Nguyen TN, Noll KM. Substrate specificities and expression patterns reflect the evolutionary divergence of maltose ABC transporters in *Thermotoga maritima*. *J Bacteriol.* 2005; 187:2002–2009. [PubMed: 15743948]
- Nanavati DM, Thirangoon K, Noll KM. Several archaeal homologs of putative oligopeptide-binding proteins encoded by *Thermotoga maritima* bind sugars. *Appl Environ Microbiol.* 2006; 72:1336–1345. [PubMed: 16461685]
- Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, et al. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature.* 1999; 399:323–329. [PubMed: 10360571]
- Novichkov PS, Rodionov DA, Stavrovskaya ED, Novichkova ES, Kazakov AE, Gelfand MS, et al. RegPredict: an integrated system for regulon inference in prokaryotes by comparative genomics approach. *Nucleic Acids Res.* 2010a; 38:W299–W307. [PubMed: 20542910]
- Novichkov PS, Laikova ON, Novichkova ES, Gelfand MS, Arkin AP, Dubchak I, Rodionov DA. RegPrecise: a database of curated genomic inferences of transcriptional regulatory interactions in prokaryotes. *Nucleic Acids Res.* 2010b; 38:D111–D118. [PubMed: 19884135]
- Osterman A, Overbeek R. Missing genes in metabolic pathways: a comparative genomics approach. *Curr Opin Chem Biol.* 2003; 7:238–251. [PubMed: 12714058]
- Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, et al. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* 2005; 33:5691–5702. [PubMed: 16214803]

- Peekhaus N, Conway T. What's for dinner?: Entner-Doudoroff metabolism in *Escherichia coli*. *J Bacteriol*. 1998; 180:3495–3502. [PubMed: 9657988]
- Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol*. 2009; 26:1641–1650. [PubMed: 19377059]
- Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. The Pfam protein families database. *Nucleic Acids Res*. 2012; 40:D290–D301. [PubMed: 22127870]
- Richard P, Hilditch S. D-galacturonic acid catabolism in microorganisms and its biotechnological relevance. *Appl Microbiol Biotechnol*. 2009; 82:597–604. [PubMed: 19159926]
- Ridley BL, O'Neill MA, Mohnen D. Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry*. 2001; 57:929–967. [PubMed: 11423142]
- Rodionov DA. Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chem Rev*. 2007; 107:3467–3497. [PubMed: 17636889]
- Rodionov DA, Gelfand MS, Hugouvieux-Cotte-Pattat N. Comparative genomics of the KdgR regulon in *Erwinia chrysanthemi* 3937 and other gamma-proteobacteria. *Microbiology*. 2004; 150:3571–3590. [PubMed: 15528647]
- Rodionov DA, Yang C, Li X, Rodionova IA, Wang Y, Obratsova AY, et al. Genomic encyclopedia of sugar utilization pathways in the *Shewanella* genus. *BMC Genomics*. 2010; 11:494. [PubMed: 20836887]
- Rodionova IA, Yang C, Li X, Kurnasov OV, Best AA, Osterman AL, Rodionov DA. Diversity and versatility of the *Thermotoga maritima* of sugar kinome. *J Bacteriol*. 2012 published ahead of print. 10.1128/JB.01136-12
- Smiley JD, Ashwell G. Uronic acid metabolism in bacteria. III. Purification and properties of D-altronic acid and D-mannonic acid dehydrases in *Escherichia coli*. *J Biol Chem*. 1960; 235:1571–1575. [PubMed: 13831814]
- Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*. 2005; 33:W244–W248. [PubMed: 15980461]
- Suresh C, Rus'd AA, Kitaoka M, Hayashi K. Evidence that the putative alpha-glucosidase of *Thermotoga maritima* MSB8 is a pNP alpha-D-glucuronopyranoside hydrolyzing alpha-glucuronidase. *FEBS Lett*. 2002; 517:159–162. [PubMed: 12062428]
- Wang Y, Zhang YH. Overexpression and simple purification of the *Thermotoga maritima* 6-phosphogluconate dehydrogenase in *Escherichia coli* and its application for NADPH regeneration. *Microb Cell Fact*. 2009; 8:30. [PubMed: 19497097]
- Yang C, Rodionov DA, Li X, Laikova ON, Gelfand MS, Zagnitko OP, et al. Comparative genomics and experimental characterization of N-acetylglucosamine utilization pathway of *Shewanella oneidensis*. *J Biol Chem*. 2006; 281:29872–29885. [PubMed: 16857666]
- Yang C, Rodionov DA, Rodionova IA, Li X, Osterman AL. Glycerate 2-kinase of *Thermotoga maritima* and genomic reconstruction of related metabolic pathways. *J Bacteriol*. 2008; 190:1773–1782. [PubMed: 18156253]
- Zhang Y, Thiele I, Weekes D, Li Z, Jaroszewski L, Ginalski K, et al. Three-dimensional structural view of the central metabolic network of *Thermotoga maritima*. *Science*. 2009; 325:1544–1549. [PubMed: 19762644]

Fig 1.

Reconstruction of the galacturonate and glucuronate catabolic pathways in *T. maritima*.

Enzymes characterized in this work are marked with asterisks. Upstream pathway components involved in polysaccharide hydrolysis and uptake are shown in a large gray box with transport routes indicated by dashed lines. The alternative biochemical route operational in *E. coli* is shown by crossed-out dotted arrow. Genes predicted to be co-regulated by the novel transcription factor UxaR in *T. maritima* are underlined. The AguA-catalysed reaction of cleavage of lateral D-glucuronate residues from the main α -xyloside chain is shown by an arrow ended by a crossline.

Fig. 2.

Pectin and hexuronate utilization gene loci and predicted UxaR regulons in *Thermotoga* spp.

A. Genomic organization of *TM0430–443* and orthologous gene clusters differ in five *Thermotoga* spp. Genes and candidate UxaR-binding sites are shown as solid arrows and red circles respectively. Orthologous genes and genes encoding components of the same transport system are shown in the same colour. Potential promoters and terminators are shown by small arrows and stems respectively. Additional functional roles predicted for genes in the orthologous loci: AldH, candidate alcohol dehydrogenase; PemA, putative pectin esterase; TRQ2_0500, sugar ABC transporter; Tpet0490, β -glucuronidase; Tpet0489-485, sugar ABC transporter; UxaXYZ, sugar ABC transporter; UxaPQM, sugar TRAP transporter.

B. Genomic organization of *TM0064–69* gene cluster is conserved in five *Thermotoga* genomes. Candidate KdgR and UxaR binding sites are shown by blue and red circles respectively.

C. Sequence logos for predicted UxaR and KdgR binding motifs according to the RegPrecise database (<http://regprecise.lbl.gov>).

Taxonomy	Microorganism	UxaE gene ID	UxaE	UxaC	UxaD	UxaB	UxaA	GntE	GntK	Gnd	KdgK	KdgA	UxaB	UxaA	Catabolic pathway variant
Spirochaetes	<i>Spirochaeta smaragdinae</i>	Spirs_3461	+	-	+	-	-	-	-	+	++	++	-	-	?
	<i>Pediococcus acidilactici</i>	KIW_09225	+	+	-	-	⊖	+	-	-	-	-	-	-	C
	<i>Lactobacillus brevis</i>	LVIS_0149	+	+	-	+	⊖	++	-	-	-	+	-	-	A, C
Spirochaetes	<i>Treponema azotonutricium</i>	TREAZ_2559	+	+	-	-	-	-	+	+	-	-	-	-	?
	<i>Spirochaeta</i> sp. Buddy	SpBuddy_2789	+	+	-	+	+	-	-	-	++	++	+	+	A, B
	<i>Paenibacillus mucilaginosus</i>	KNP414_02589	+	+	-	-	-	+	+	+	+	+	+	+	A
Firmicutes-Bacilli	<i>Oceanobacillus iheyensis</i>	OB0366	+	+	+	-	-	+	+	+	+	+	+	+	A
	<i>Halanaerobium hydrogeniformans</i>	Halsa_0911	+	+	+	+	+	+	+	+	+	+	+	+	A, B
	<i>Tepidanaerobacter</i> sp.	TepRe1_1581	+	-	-	-	-	+	+	+	+	+	+	+	?
Firmicutes-Clostridia	<i>Petrotoga mobilis</i>	Pmob_0856	+	+	+	-	⊖	+	+	+	+	+	+	+	C
	<i>Coralimargarita akajimensis</i>	Caka_0459	+	+	+	-	-	-	-	+	++	++	+	+	A, B
	<i>Terriglobus saanensis</i>	AcIPR4_0407	+	+	+	-	-	⊖	+	+	+	+	+	+	C
Acidobacteria	<i>Acidobacterium</i> sp.	Acix9_2071	+	+	+	-	-	⊖	+	+	+	+	+	+	C
	<i>Acidobacterium capsulatum</i>	ACP_2681	+	+	+	-	-	⊖	+	+	+	+	+	+	C
	<i>Solibacter usitatus</i>	Acid_7772	+	+	+	-	-	-	+	+	+	+	+	+	A
Spirochaetes	<i>Spirochaeta thermophila</i>	Spith_0401	+	+	+	-	+	+	+	+	+	+	+	+	A
	<i>Spirochaeta caldaria</i>	Spica_0065	+	+	+	-	-	-	+	+	+	+	+	+	A
	<i>Pedobacter saltans</i>	Pedsa_2015	+	+	+	-	⊖	+	+	+	+	+	+	+	A, B, C
Bacteroidetes	<i>Thermoanaerobacter xylanolyticum</i>	Thexy_2043	+	+	-	+	⊖	+	+	+	+	+	+	+	A, B, C
	<i>T. thermosaccharolyticum</i>	Tthe_2391	+	+	-	+	⊖	+	+	+	+	+	+	+	A, B, C
	<i>Thermoanaerobacter italicus</i>	Thit_1746	+	++	-	++	⊖	+	++	+	+	+	+	+	A, C
Firmicutes-Clostridia	<i>T. tengcongensis</i>	TTE1944	+	+	-	-	⊖	+	+	+	+	+	+	+	C
	<i>Dictyoglomus turgidum</i>	Dtur_0194	+	⊖	-	-	⊖	+	+	+	+	+	+	+	C
	<i>Dictyoglomus thermophilum</i>	DICTH_1923	+	⊖	-	-	⊖	+	+	+	+	+	+	+	C
Dictyoglomi	<i>Halorhabdus utahensis</i>	Huta_0157	+	++	-	-	⊖	+	+	+	+	+	+	+	A, C
	<i>Planctomyces brasiliensis</i>	Plabr_3924	+	⊖	-	-	-	-	-	-	-	-	-	-	?
	<i>Brachybacterium faecium</i>	Bfae_07190	+	+	-	-	-	-	-	-	-	-	-	-	?
Various phyla	<i>Rhodothermus marinus</i>	Rmar_1528	+	+	-	-	⊖	+	+	+	+	+	+	+	?
	<i>Truepera radiovictrix</i>	Trad_0663	+	+	+	-	+	⊖	+	+	+	+	+	+	C
	<i>Thermotoga lettingae</i>	Tlet_0825	+	+	+	-	+	+	+	+	+	+	+	+	C
Thermotogae	<i>Thermotoga thermarum</i>	Theth_1083	+	+	+	+	⊖	+	+	+	+	+	+	+	A, C
	<i>Thermotoga neapolitana</i>	CTN_0232	+	+	+	+	+	+	+	+	+	+	+	+	A, C
	<i>Thermotoga maritima</i>	TM0440	+	+	+	+	+	+	+	+	+	+	+	+	A, C

Fig. 3.

Distribution and genomic context of UxaE and other hexuronate utilization genes in prokaryotic genomes. The phylogentic cladogram of UxaE proteins was constructed using the maximum likelihood approach and presented as a phylogram with support values in Fig. S2. Genomic identifiers (locus tags) of genes encoding UxaE proteins are listed. Three *Thermotoga* spp. with identical distribution and context of the gal catabolic genes as in *Thermotoga maritima*, namely *Thermotoga* sp. RQ-2, *Thermotoga petrophila* and *Thermotoga naphthophilila*, are not displayed. Presence and absence of orthologues of the *T. maritima* genes are shown by '+' and '-' respectively. Background colours show colocalization of genes in the chromosome. Two or three paralogues found in the same genome are shown by '++' or '+++'. Functional roles of the analysed enzymes are: UxaE, D-TagA/D-FruA epimerase; UxaC, hexuronate isomerase; UxuB, D-ManA dehydrogenase; UxaD, D-FruA reductase; GntE, uncharacterized enzyme; GntK, D-gluconate kinase; Gnd, phosphogluconate dehydrogenase; UxuA, D-ManA dehydratase; KdgK, KDG kinase; KdgA, KDG-6-P aldolase; UxaB, D-AltA dehydrogenase; UxaA, D-AltA hydrolase. Non-orthologous displacements of the *T. maritima* UxaC, GntK and GntE enzymes in some genomes are circled. UxaC is replaced by an orthologue of the hexuronate isomerase BH0493. The FGGY family GntK enzyme is replaced by a gluconokinase from the shikimate kinase family (as GntK in *E. coli*). GntE is possibly replaced by an uncharacterized dehydrogenase from the LdhA family. The last column indicates the existence of one or several variants of the hexuronate catabolic pathway (see *Discussion* and Fig. 9).

Fig. 4.

Biochemical characterization of recombinant proteins *tm_UxuB* and *tm_UxaD*.

A. D-ManA dehydrogenase activity of *tm_UxuB*/TM0068 was determined by monitoring the increase in absorbance at 340 nm due to conversion of NAD⁺ to NADH.

B. D-FruA reductase activity of *tm_UxaD*/TM0441 was determined by monitoring the decrease in absorbance at 340 nm due to conversion of NADPH to NADP⁺.

Fig. 5. Characterization of the product of *tm_UxaD*-catalysed reduction of D-FruA by GC-MS. The initial D-FruA/D-GlcA mixture (dotted line) was incubated overnight with the recombinant *tm_UxaD*/TM0441 protein and NADPH, resulting in the substantial decrease of two D-FruA-specific peaks and the appearance of a new peak (solid line). Retention time of this novel peak (10.75 min) on the total ion count chromatogram is identical to retention time of D-ManA standard.

Fig. 6.

Characterization of the product of *tm_UxaE*-catalysed epimerization of D-FruA by GC-MS. The initial D-FruA/D-GlcA mixture (dotted line) was incubated overnight with the recombinant *tm_UxaE*/TM0440 protein, showing the decrease of two D-FruA-specific peaks and the appearance of two new peaks with retention times 10.39 and 10.51 min (solid line). These two peaks were associated with the D-TagA product based on comparison with the chromatograms and electron ionization spectra obtained for *tm_UxaC*-catalysed conversion of D-GalA/D-GlcA to D-TagA/D-FruA (Figs S3 and S5).

Fig. 7.

Biochemical characterization of recombinant *tm_UxaE* protein.

A. D-FruA epimerase activity of *tm_UxaE*/TM0440 was determined by the coupling colorimetric assay with NADH-dependent D-TagA reductase from *E. coli* (*ec_UxaB*). B. Temperature dependence of D-TagA epimerase activity of *tm_UxaE*/TM0440 was determined by the coupling colorimetric assay with the thermostable enzyme NADH-dependent D-FruA reductase from *T. maritima* (*tm_UxuB*/TM0068). The assay mixture was pre-incubated during 20 min at each studied temperature; then the *T. maritima* enzymes were added to the mixture while keeping the temperature constant. The oxidation of NADH to NAD⁺ was monitored at 340 nm.

Fig. 8.
Proposed mechanism of reaction for D-TagA/D-FruA epimerase UxaE.

Fig. 9.

Distribution of three proposed variants of UxaE-involving pathways for hexuronate utilization in Bacteria.

A. Reconstructed metabolic network of hexuronate utilization in Bacteria. A novel reaction proposed in this work for uncharacterized protein GntE is shown by a dotted line and marked with a question mark.

B. The Venn diagram of distribution of three catabolic pathway variants in UxaE-encoding bacterial genomes.

Table 1

Predicted functional assignments in pectin and D-galacturonate utilization pathway in *T. maritima*.

Gene	ID	Suggested function ^a	Evidence ^b	Experimental verification ^c
<i>pelA</i>	<i>TM0433</i>	Pectate lyase (EC 4.2.2.2) [PL1]	E, H	Kluskens and colleagues (2003)
<i>aguEFG</i>	<i>TM0430-31-32</i>	Digalacturonate-specific ABC transporter	E, GC	Nanavati and colleagues (2006)
<i>pelB</i>	<i>TM0437</i>	Exo-polygalacturonase (EC 3.2.1.15) [GH28]	E, H	Kluskens and colleagues (2005)
<i>agu4A</i>	<i>TM0434</i>	α -Glucuronidase (EC 3.2.1.139) [GH4]	E, H	Suresh and colleagues (2002)
<i>gnd</i>	<i>TM0438</i>	6-Phosphogluconate dehydrogenase (EC 1.1.1.44)	E, H	Wang and Zhang (2009)
<i>uxaR</i>	<i>TM0439</i>	Hexuronate catabolism regulator, GntR family	E, GC	Unpublished
<i>uxaE</i>	<i>TM0440</i>	D-TagA epimerase	E, GC	This work
<i>uxaD</i>	<i>TM0441</i>	D-FruA reductase (EC 1.1.1.57)	E, H, GC	This work
<i>gntE</i>	<i>TM0442</i>	Predicted D-mannate utilization enzyme	–	–
<i>gntK</i>	<i>TM0443</i>	D-gluconate kinase (EC 2.7.1.12)	E, H	Unpublished
<i>uxaC</i>	<i>TM0064</i>	Hexuronate isomerase (EC 5.3.1.12)	E, H, GC	This work
<i>kdgR</i>	<i>TM0065</i>	Putative KDG-responsive repressor, IclR family	H, GC	–
<i>kdgA</i>	<i>TM0066</i>	KDG-6-P aldolase (EC 4.1.2.14)	E, H, GC	Griffiths and colleagues (2002)
<i>kdgK</i>	<i>TM0067</i>	KDG kinase (EC 2.7.1.45)	H, GC	–
<i>uxuB</i>	<i>TM0068</i>	D-mannate dehydrogenase (EC 1.1.1.57)	E, H, GC	This work
<i>uxuA</i>	<i>TM0069</i>	D-mannate dehydratase (EC 4.2.1.8)	H, GC	–

^aThe EC enzyme classification numbers and the CaZy carbohydrate-active enzyme families are indicated in parentheses and square brackets respectively.

^bEvidence codes: E, experimental; H, homology; GC, genomic context.

^cUnpublished experimental confirmation of functions of UxaR and GntK will be presented elsewhere.

Table 2Kinetic parameters of hexuronic acid catabolic enzymes from *T. maritima*.

Enzyme	Activity	Substrate	K_m (mM)	k_{cat} (s ⁻¹)	k_{cat}/K_m (M ⁻¹ s ⁻¹)
UxaC	D-GalA/D-TagA isomerase	D-GalA	3.5 ± 0.4	3.8 ± 0.35	1090
	D-GlcA/D-FruA isomerase	D-GlcA	2.6 ± 0.7	0.35 ± 0.04	140
UxaE	D-TagA/D-FruA epimerase	D-FruA	0.68 ± 0.1	6.5	9100
UxuB	D-ManA dehydrogenase	D-ManA	0.63 ± 0.05	4.2 ± 0.1	6670
UxaD	D-FruA reductase	D-FruA	10 ± 3.0	0.05	5
	5-Keto-D-gluconate reductase	5-Keto-D-gluconate	7.4 ± 1.6	0.06	9