



OPEN

Chemical structure and genetic organization of the *E. coli* O6:K15 capsular polysaccharide

Hugo F. Azurmendi¹, Vamsee Veeramachineni¹, Stephen Freese², Flora Lichaa¹, Darón I. Freedberg¹ & Willie F. Vann¹✉

Capsular polysaccharides are important virulence factors in pathogenic bacteria. Characterizing the structural components and biosynthetic pathways for these polysaccharides is key to our ability to design vaccines and other preventative therapies that target encapsulated pathogens. Many gram-negative pathogens such as *Neisseria meningitidis* and *Escherichia coli* express acidic capsules. The *E. coli* K15 serotype has been identified as both an enterotoxigenic and uropathogenic pathogen. Despite its relevance as a disease-causing serotype, the associated capsular polysaccharide remains poorly characterized. We describe in this report the chemical structure of the K15 polysaccharide, based on chemical analysis and nuclear magnetic resonance (NMR) data. The repeating structure of the K15 polysaccharide consists of 4)- α -GlcNAc-(1 \rightarrow 5)- α -KDOp-(2 \rightarrow partially O-acetylated at 3-hydroxyl of GlcNAc. We also report, the organization of the gene cluster responsible for capsule biosynthesis. We identify genes in this cluster that potentially encode an O-acetyltransferase, an N-acetylglucosamine transferase, and a KDO transferase consistent with the structure we report.

Capsular polysaccharides are major virulence factors of pathogenic *Escherichia coli*. These surface glycans form a coat exterior to most other surface structures such as lipopolysaccharides, protecting the bacteria from a multitude of environmental stress factors and serving as a barrier to the host immune responses¹. The presence of specific polysaccharide capsules on the surfaces of some strains of bacteria that cause disease, make capsules, useful serological targets for monitoring and prevention of infection. *Escherichia coli* has more than 80 reported serologically distinct capsular polysaccharides, referred to as K-antigens². Capsules in *E. coli* have been classified into 4 groups based on organization of the capsule synthesizing genes, mechanism of biosynthesis, and mode of transport across the membranes³. Groups 1 and 4 capsules use the Wzy-dependent pathway, while groups 2 and 3 make use of the ABC transporter dependent pathway of assembly and export.

Most capsules of pathogenic *E. coli* contain anionic glycan residues like hexuronic acid, N-acetyl neuraminic acid (NeuNAc), 2-keto-3-deoxy-D-mannooctulonic acid (KDO), and phosphodiester⁴. Among these negatively charged components, KDO has only been identified in group 2 capsules. In capsular polysaccharide structures, KDO is usually present in a β -pyranosidic (β -KDOp, in *E. coli* K12, K16, K20, K23, etc.) and β -furanosidic forms (β -KDOf, in *E. coli* K74 and K95)⁵⁻⁹. The only known examples of KDO as α -KDOp in capsular polysaccharides were reported for *E. coli* K6 (LP1092) and K16¹⁰⁻¹³. On the other hand, α -KDOp is widely reported as part of the lipid A core component of the lipopolysaccharide (LPS). The genetic organization of several gene clusters encoding the biosynthesis of capsular polysaccharides containing KDO with putative glycosyltransferases have been described¹⁴. However, a description of the genes encoding the specific glycosyltransferases and O-acetyltransferase for synthesis of the K15 polysaccharide has not been reported. Therefore, determining the structure of these polysaccharides as well as the genetic organization of the biosynthetic machinery, is crucial to understanding infection and developing of effective treatments.

E. coli K15 is an enterotoxigenic K antigen type originally isolated from children with diarrhea and occurring at high frequency in combination with the O6 antigen and a mannose resistant hemagglutinin^{15,16}. Orskov et al. demonstrated, in a large study of enteropathogenic *E. coli* strains from adults and children from widespread geographic locations, that some O and K serotypes frequently occurred together^{17,18}. They suggested that these strains represent clones which have adapted to growth in the small intestine. A structure was reported for the *E. coli* K15 polysaccharide without experimental details of how this structure was determined nor the source the polysaccharide⁷. Significantly, this reported structure differs from our findings. In this study, we describe

¹Laboratory of Bacterial Polysaccharides, Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, MD 20993, USA. ²Affinivax, 650 East Kendall St, Cambridge, MA 02138, USA. ✉email: willie.vann@fda.hhs.gov

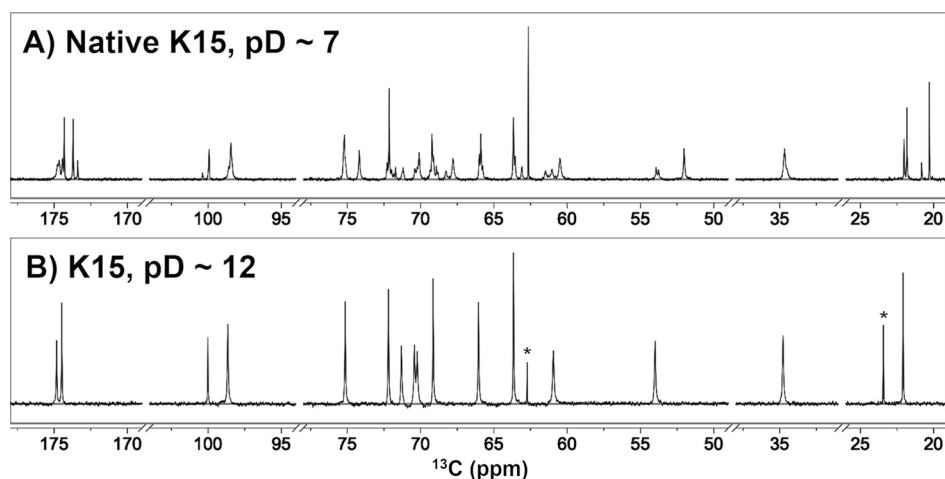


Figure 1. ^{13}C 1D NMR of the K15 polysaccharide at 50 °C. (A) Native K15 at pD ~ 7. (B) K15 PS after raising the pD to ~ 12. A total of 16 signals can be identified belonging to the K15 PS in panel (B); the ‘*’ marks indicate narrow trace signals from K15 decomposition (acetate and glycerol).

structural analysis of capsular polysaccharide isolated from *E. coli* str. F8316/41 (O6:K15:H16). We also report the partial characterization of the gene cluster responsible for biosynthesis of the K15 capsular polysaccharide in the same strain of *E. coli* K15. This gene cluster is very similar to that of *E. coli* K15 strain 536¹⁹ whose polysaccharide structure has not been reported.

Results and discussion

Composition of the K15 polysaccharide. The composition of the K15 capsular polysaccharide was determined by a combination of chemical degradation methods and analysis of its one-dimensional carbon nuclear magnetic resonance (1D ^{13}C -NMR) spectrum. The polysaccharide is composed of two monosaccharides, *N*-acetylglucosamine (GlcNAc) and KDO. *N*-acetylglucosamine and KDO were detected exclusively in acid methanolsates by gas-chromatography mass-spectrometry (GC-MS) as trimethylsilyl (TMS) derivatives. The identity of the hexosamine in the polysaccharide was confirmed by ninhydrin degradation of an acid hydrolysate. The product of the ninhydrin degradation was identified chromatographically as arabinose, the expected degradation product for glucosamine²⁰. KDO was identified by acid hydrolysis of the carboxyl reduced polysaccharide, conversion to alditol acetates and analysis by GC-MS for characteristic fragmentation patterns. These composition results are in agreement with the chemical shift location and number of carbon signals observed in 1-D ^{13}C -NMR spectra (Fig. 1). Treatment of the native polysaccharide (Fig. 1A) with dilute sodium hydroxide results in loss of methyl and carbonyl carbon signals, simplifying the spectrum (Fig. 1B). The alkali treated polysaccharide has two chemical shifts in the anomeric region at 98.5 and 100 ppm, two carbonyl chemical shifts at 174.46 and 174.83 ppm and a single methyl chemical shift at 22.09 ppm. The loss of some methyl and carbonyl carbon signals under alkaline conditions with a corresponding simplification of the spectrum is characteristic of an *O*-acetylated polysaccharide²¹.

The presence of only two resonances of similar intensities in the anomeric region of the ^{13}C -NMR spectra of the de-*O*-acetylated polysaccharide suggests a disaccharide repeat unit. The reduction of the number of signals after de-*O*-acetylation to 16 and their chemical shift positions are expected for a disaccharide repeat unit composed of an *N*-acetylhexosamine and a deoxy acidic sugar with eight carbons, i.e., KDO.

An oligosaccharide consisting of KDO and GlcNAc was isolated by selective cleavage at the acid labile KDO linkage with 1% acetic acid. This oligosaccharide eluted on a BioGel P2 gel filtration as a disaccharide, which is in agreement with the suggestion of a disaccharide repeat unit from the NMR evidence above. Methylation analysis of the purified oligosaccharide yielded only 3,4,6-*O*-methyl-*N*-acetylglucosamine, as expected for a terminal GlcNAc, indicating that the repeat unit consists of a disaccharide of GlcNAc and KDO.

Substitution of repeat unit. The substitution pattern of the monosaccharides in the polysaccharide repeat unit was determined by two dissimilar methylation analysis experiments, due to the difference in lability of GlcNAc and KDO. The GlcNAc linkage was determined by permethylation of the polysaccharide and subsequent analysis by GC-MS of the resulting alditol acetate, which yielded only 3,6-*O*-methyl-*N*-acetylglucosamine, suggesting that this glycan is 4-substituted. A permethylated KDO residue was not conveniently detected during this analysis, thus another approach was taken. The permethylated K15 was treated with LiBD_4 to reduce the carboxylate group, and subsequently hydrolyzed with TFA and converted to alditol acetates before GC-MS. The major fraction detected in this analysis corresponded to a 2,5 linked KDO suggesting that the KDO is 5-substituted (Fig. 2). This was further confirmed by the alkali lability of the periodate oxidized polysaccharide. The K15 polysaccharide is oxidized by periodate with concomitant loss of KDO but remains polymeric. The oxidized product is depolymerized by mild alkali, suggesting a β -elimination reaction. Depolymerization was easily fol-

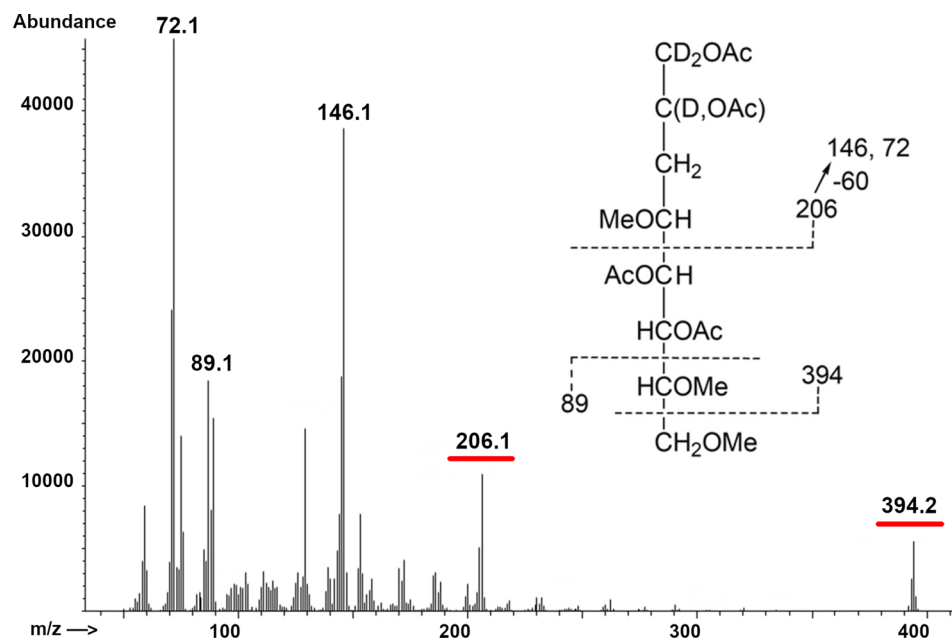


Figure 2. Mass spectrum and fragmentation profile of 2,5 linked KDO labeled peak identified in reduced permethylated K15 polysaccharide. Carboxylate reduced permethylated K15 polysaccharide was hydrolyzed, converted to alditol acetates and analyzed by GC-MS.

lowed by TLC, which showed the disappearance of immobile polysaccharide and the appearance of a component with $R_{f\text{GlcNAc}} = 0.75$. Thus, the chemical analysis indicates that the repeat unit of the K15 CPS is a disaccharide of 4-substituted GlcNAc and 5-substituted KDO.

The analysis above was confirmed by NMR experiments. We used the following process to confirm the composition and substitution of the repeat unit. The HSQC-TOCSY^{22,23} experiment produces signals from single-bonded CHs correlated to ¹H signals from Hs belonging to the same ring as the CH. Since there are multiple carbon atoms in each ring, this experiment yields redundant information that allows us to confirm ¹H assignments. An overlay of the ¹H-¹³C HSQC and HSQC-TOCSY (Fig. 3A), allows assignment of the ¹H and ¹³C chemical shifts in the KDO and GlcNAc rings, as follows. In the HSQC-TOCSY, cross-peak intensities originating from a given ¹H decrease as the number of bonds between the ¹Hs increases, thus suggesting tentative assignments based on relative intensities. For instance, all ¹H resonances in the GlcNAc ring can be identified and assigned tentatively from the ¹³C slice corresponding to G-C1 (Fig. 3A). The anomeric G1 is found at ~ (98.6, 5.07) ppm in ¹³C and ¹H, respectively (Table 1). From G1 it is possible to use the HSQC-TOCSY to provisionally assign all ¹H resonances through G5. Signals from H5-C5 and H4-C4 have lower intensity than those from H3-C3 and H2-C2. Similarly, the characteristic signals for the equatorial and axial H3's in KDO, H3e,a, respectively in KDO^{24,25}, allow tentative assignment of H4 and H5 at the C3 slice (~ 34.7 ppm).

To confirm the initial assignments and complete the remaining ones, we added to our analysis a low-pass filtered HMBC^{26,27}, which produces only long-range ¹H-¹³C correlations. Figure 3B shows an overlay of the HSQC (red) with the HMBC (blue) spectra. The presence of peaks depends on both, the ¹H-¹³C couplings between each specific pair of atoms and multiple nuclear relaxation rates. Consequently, not all potential signals are observed. For instance, there is no cross-peak for (GH1- GC2) but there is one between a related pair of nuclei, GH2 and GC1. The HMBC also provided evidence of the G-K linkage: a cross-peak between G1 and K5, indicated a G1-K5 linkage between the glycan residues. In addition, the ¹³C chemical shift assignments of K2 ($\delta_C = 100.0$ ppm) and N-acetyl carbonyl group ($\delta_C = 174.5$ ppm) are confirmed through the cross-peaks K(H3e, C2) and G(CH3, C7), respectively. This indicates that the other carbonyl peak ($\delta_C = 174.8$ ppm) corresponds to K1.

The analysis continued with the ¹H assignments of the H3a and H3e of KDO in the K15 polysaccharide using a ¹H-¹³C high-resolution HSQC at 70 °C, which allowed determination of ¹H-¹H couplings between K-H3a/e and K-H4 (Fig. 4). H4 is axial in KDO α , the resulting $^3J_{H3aH4}$ is large (H4-C4-C3-H3a torsion of ~ 174°) and similar to $^2J_{H3aH3e}$ in magnitude, while $^3J_{H3eH4}$ is small (H4-C4-C3-H3e torsion of ~ 56°) compared to $^2J_{H3aH3e}$. Thus, K-H3e displays a doublet at 2.10 ppm ($^3J_{HH} = 12.6$ Hz) while K-H3a results in a triplet at 1.89 ppm ($^3J_{HH} = 13.0$ Hz).

Multiple NMR parameters are indicative of the anomeric configuration of pyranose KDO (KDO α , Fig. 5A), most notably the δ_H of the axial and equatorial H3 (H3a and H3e, respectively)²⁸, and the C-H three-bond coupling ($^3J_{CH}$) between H3a and C1²⁹. From δ_H analysis, it has been observed that a reliable criterion for distinguishing α KDO and β KDO is the signal separation between δ_{H3a} and δ_{H3e} (or $\Delta(\delta_{H3a}, \delta_{H3e})$)¹³; for α KDO $\Delta(\delta_{H3a}, \delta_{H3e}) < 0.4$ ppm, while for β KDO $\Delta(\delta_{H3a}, \delta_{H3e}) > 0.4$ ppm (Supplementary Table S1). For K15 $\Delta(\delta_{H3a}, \delta_{H3e}) = 0.21$ ppm, further validating that KDO is in the α -configuration in K15. Based on a representative survey (Supplementary Table S1), we propose that the same conclusion can be reached using only the δ_{H3e} , as for all reported values it is verified that $\delta_{H3e}(\alpha\text{KDO}) < 2.3$ ppm and $\delta_{H3e}(\beta\text{KDO}) > 2.3$ ppm. For K15 the δ_{H3e} signal is ~ 2.1 ppm, well below the 2.3 ppm threshold, further confirming the configuration as α KDO.

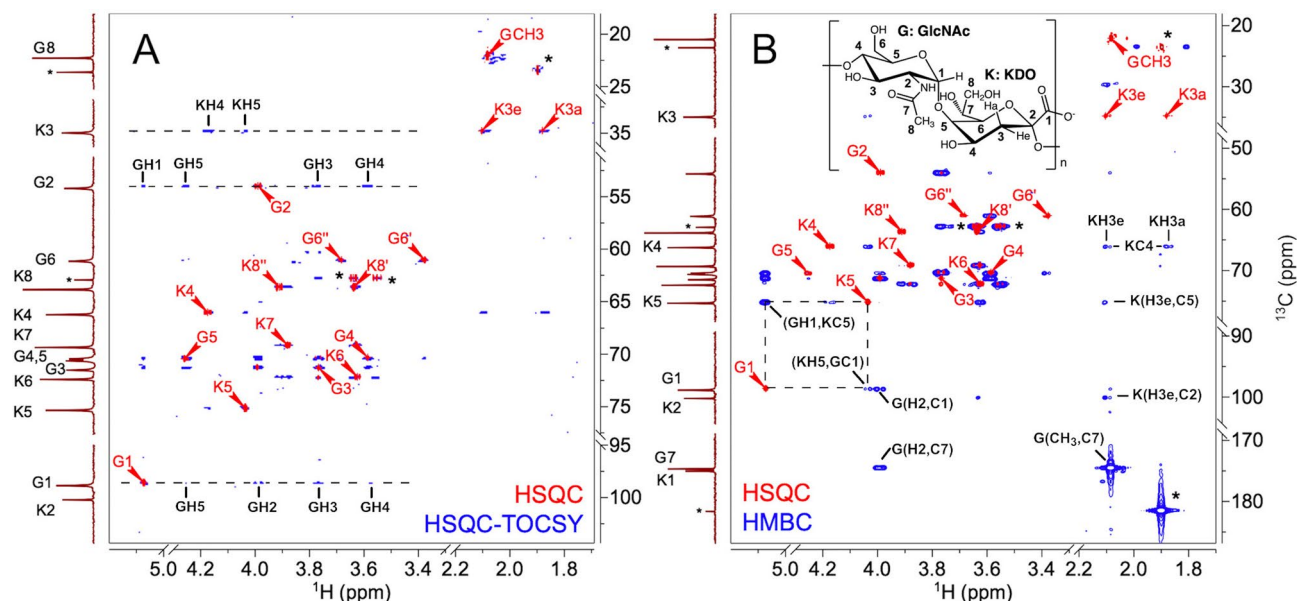


Figure 3. NMR assignments of the K15 CPS at pD ~ 12 and T = 50 °C. For reference the same HSQC is overlaid in both panels (red) with the C-H peak assignments, and a ^{13}C 1D spectrum is shown as vertical trace. **(A)** HSQC-TOCYSY spectrum acquired with a 60 ms mixing time (blue), showing the ^1H - ^1H spin coupling at different ^{13}C chemical shifts. Horizontal dashed lines highlight examples of ^1H - ^1H spin systems: at GlcNAc C1 (98.55 ppm, indicated by G1), GlcNAc C2 (53.97 ppm, indicated by G2) and KDO C3 (34.75 ppm, indicated by K3) ^{13}C chemical shifts. **(B)** HMBC spectrum (blue) for long-range connectivity was assigned by NMR. The G1-K5 HMBC interglycosidic correlation confirms the linkage in the repeating disaccharide and is indicated by dashed lines. The * marks indicate trace signals from K15 decomposition (acetate and glycerol). (A high-resolution image of panel A is provided as Figure S1).

Bond	K15(-OAc)		K15(G3+OAc)	
	δ_{C} (ppm)	δ_{H} (ppm)	δ_{C} (ppm)	δ_{H} (ppm)
G1	98.55	5.070	98.46	5.127
G2	53.97	3.988	52.05	4.219
G3	71.25	3.769	74.20	5.132
G4	70.25	3.589	67.84	3.796
G5	70.41	4.257	70.14	4.369
G6'	61.1	3.387	60.5	3.400
G6''	61.1	3.692	60.5	3.708
CH ₃ (NAc)	22.09	2.083	21.88	2.027
C(NAc)	174.46	–	174.36	–
CH ₃ (OAc)	–	–	20.34	2.103
CO(OAc)	–	–	173.76	–
K1	174.83	–	174.81	–
K2	100.01	–	99.95	–
K3e	34.75	1.887	34.67	1.946
K3a	34.75	2.098	34.67	2.125
K4	65.92	4.167	66.02	4.187
K5	75.1	4.037	75.25	4.079
K6	72.18	3.612	72.30	3.592
K7	69.2	3.891	69.2	3.891
K8'	63.6	3.636	63.7	3.576
K8''	63.6	3.915	63.7	3.923

Table 1. ^1H and ^{13}C NMR assignments for K15 \pm (OAc).

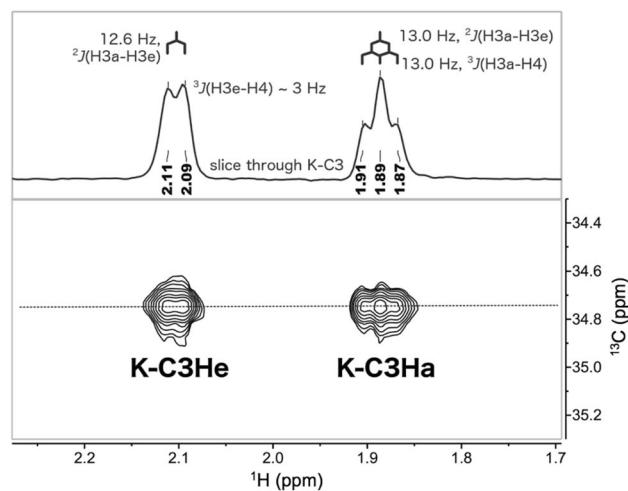


Figure 4. High resolution ^1H - ^{13}C HSQC of the K15 PS at 70 °C, showing the couplings to K-H4 from K-H3a and K-H3e. The two different $^3J_{\text{HH}}$ couplings to K-H4 in KDO allowed the assignment δH at 2.10 ppm as the equatorial H3 (K-H3e) and the δH at 1.89 ppm as the axial H3 (K-H3a).

The $^3J_{\text{CH}}$ criteria is based on the fact that the H3a-C3-C2-C1 torsion for the β -configuration is $\sim 166^\circ$, which results in a $^3J_{\text{CH}}$ coupling of ~ 6 Hz. For the α -configuration the same torsion is $\sim 60^\circ$, yielding a $\text{K}(\text{H3a-C1})$ $^3J_{\text{CH}}$ coupling of < 1 Hz (Fig. 5B). The torsion between H3e and C1 is about the same ($\sim 60^\circ$) for both α - and β -KDO, thus, the $^3J_{\text{CH}}$ value for the H3e-C3-C2-C1 torsion is not informative for establishing the α or β anomeric configuration in KDO. For the K15 PS, the KC1 (174.8 ppm) and KC2 (100.0 ppm) ^{13}C 1D signals (Fig. 5C) are broad preventing a precise measurement of the couplings, however, these peaks are similar to those in Fig. 5B arising from $\alpha\text{KDO-C1}$ (176.8 ppm) and C2 (96.4 ppm), and dissimilar to those arising from $\beta\text{KDO-C1}$ (175.5 ppm) and C2 (97.4 ppm) signals, providing independent confirmation of the presence of α -KDO in the K15 PS.

The δ_{H} of ~ 5.1 ppm indicates that GlcNAc is in the α -configuration, as δ_{H} ranges between 4.4 and 4.8 ppm for β -GlcNAc-H1^{30,31}. The measured one-bond coupling ($^1J_{\text{CH}}$) for GlcNAc-C1H1 is ~ 172 Hz (Fig. 5C), also indicating that the GlcNAc ring is in the α configuration, expected $^1J_{\text{CH}}$ values for β -GlcNAc-C1H1 are about 161 Hz³².

The native (*O*-acetylated) PS shows several NMR signals that cannot be directly identified by comparison between the native and de-*O*-acetylated NMR spectra (Supplementary Fig. S2A). Signals for the native PS, were completely assigned running the same set of experiments as for the de-*O*-acetylated sample and indicated that G-C3 is *O*-acetylated. *O*-acetylation deshields the G-H3 signal by ~ 1.35 ppm, by far the most pronounced change in δ_{H} for K15 resulting from *O*-acetylation. The δ_{H} at ~ 5.1 ppm of G-H3-OAc is typical for *O*-acetylated positions³³.

Table 1 contains the ^1H and ^{13}C NMR assignments for de-*O*-acetylated K15, and the major population of native K15. Except for the *O*-acetylated G3 and its immediate neighbors, few chemical shifts differ for the native and de-*O*-acetylated forms.

To obtain an approximate *O*-acetylation ratio in the K15 CPS, we ran a long 1D inverse-gated ^{13}C experiment to quantitate, by integration, isolated NMR signals of the *O*-acetylated PS and compare it with signals comprising total PS. Supplementary Fig. S2B shows this comparison, illustrating that $\gtrsim 60\%$ of GlcNAc residues are *O*-acetylated at the C3 position. Altogether, the data indicate that the K15 repeating unit consists of the structure in Scheme 1.

Sequencing the *E. coli* strain F8316/41(O6:K15:H16) and capsule gene cluster identification.

The genetic organization of the capsular gene cluster (in pathogenic island V) of *E. coli* strain 536 (O6:K15:H31) was previously reported to be a combination of structural features from group 2 and 3 capsule determinants¹⁹. However, the lack of insight into the capsule specific genes in region 2 and the unusual organization of the strain 536 gene cluster, prompted us to study the capsular gene cluster from *E. coli* strain F8316/41 (O6:K15:H16), from which the capsular polysaccharide was isolated for determining the structure. The chromosomal DNA isolated from *E. coli* strain F8316/41 was sequenced using single-molecule real-time (SMRT) sequencing technology³⁴ yielding five contigs (GenBank accession numbers: JAADZB010000001-JAADZB010000005). The annotated DNA contigs were screened for gene markers reported to exist in proximity to Group 2 and Group 3 capsular gene clusters of *E. coli* that use the ABC transporter pathway, namely *pgk* (gene encoding phosphoglycerate kinase), *pheV* (gene encoding phenylalanine tRNA) and *serA* (gene encoding phosphoglycerate dehydrogenase)³⁵⁻³⁷. All these gene markers along with a second 22 bp truncated copy of *pheV* (*pheV'*) were identified in the second contig (GenBank accession numbers are provided in Supplementary Table S2). We also observed that 20 kb of DNA downstream to *pheV'*, contained open reading frames (ORFs) that closely resembled K15 capsular gene cluster from *E. coli* 536¹⁹ (GenBank accession number: AJ617685) with a few variations. Given the relatively high error rate of long-read SMRT platforms we amplified 20 specific areas of interest within the gene cluster. Sanger sequencing of the amplified DNA segments identified two misreads

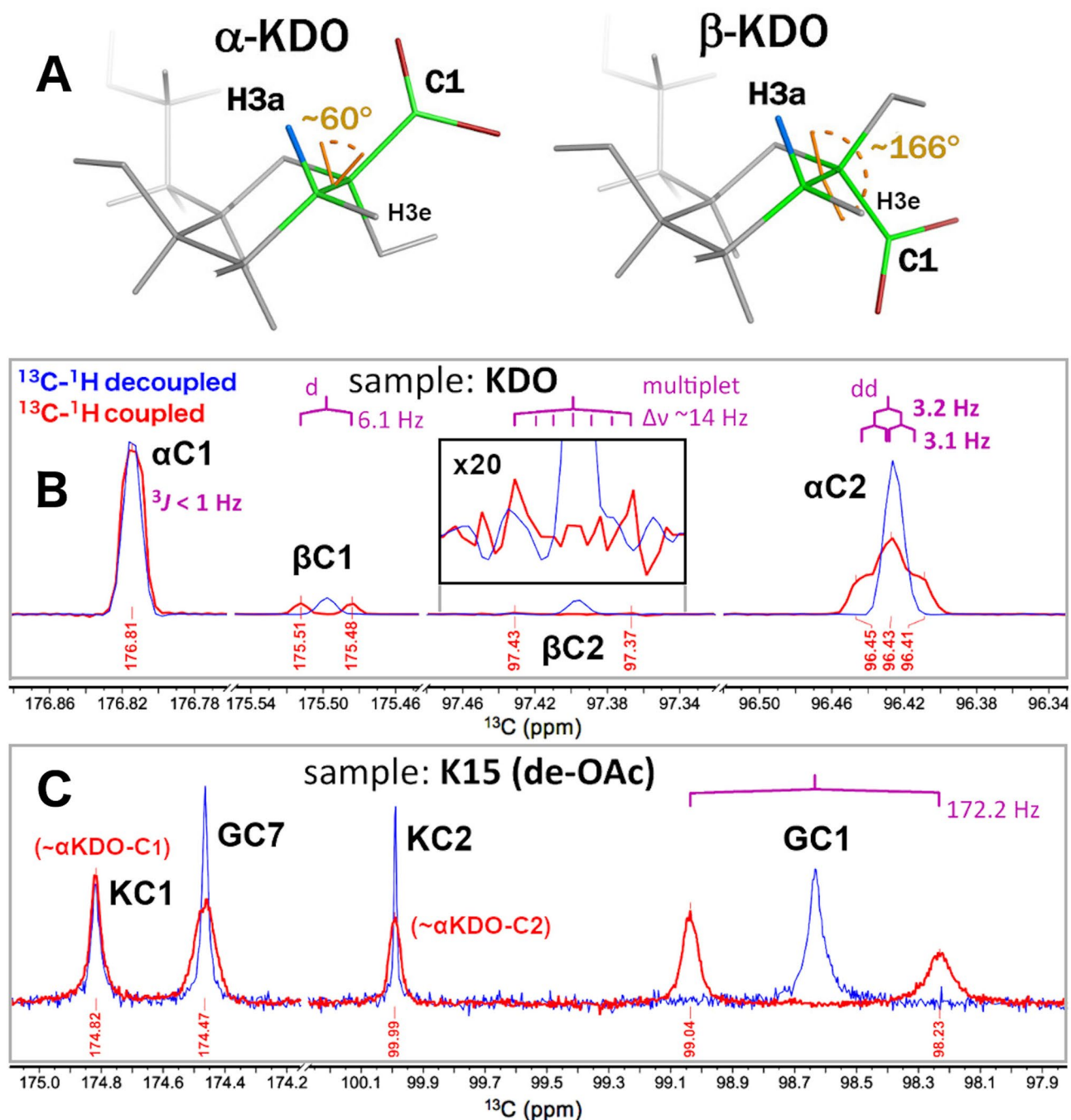
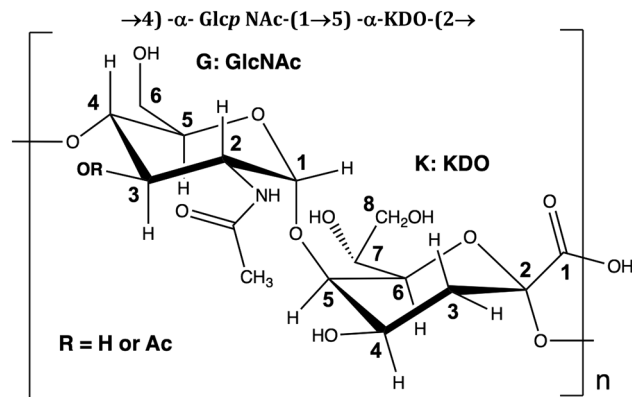


Figure 5. Comparison of the ^{13}C (^1H coupled and decoupled) α - and β -anomers of KDO monomer signals, to those of the K15 capsular polysaccharide signals for determination of the anomeric configuration of KDO in K15. (A) Schemes showing the H3a to C1 three-bond torsion in α - and β -KDO, respectively. (B, C) ^{13}C ^1H -coupled (red) and ^{13}C ^1H -decoupled spectra of KDO (B) and K15 (C), demonstrating that α -KDO is the K15 constituent. Figure 5A was created with PyMol 2.4 (www.schrodinger.com/pymol); the coordinates for α/β -KDO were obtained from CarbBuilder (www.organ.su.se/gw/doku.php?id=carbbuilder).

in the SMRT sequencing data. Correcting these misreads resulted in a capsule gene cluster which had 99.6% sequence identity with the *E. coli* 536 gene cluster¹⁹.

The K15 capsular gene cluster from *E. coli* F8316/41 (GenBank accession number: NDK77120.1-NDK77133.1), when screened against the nucleotide collection (nt) database using BLASTN identified 13 different strains of *E. coli* carrying a similar gene cluster in their genome³⁸. Of these, genomes of 7 strains were found to be annotated (NCTC11105, FORC_031, FMU073332, 536, 743, MRY15-131 and MRY15-117) and were further analyzed and compared against F8316/41. While all the 7 strains were enterobacterial, MRY15-131 and MRY15-117 were isolated from *Bos taurus*. The capsular gene cluster from all the 8 strains compared earlier belongs to a composite, *pheV*-associated pathogenicity island (PAI). A PAI represents a mosaic-like gene structure with



Scheme 1. Chemical structure of the disaccharide repeat unit of the K15 polysaccharide.

multiple functional and fragmented mobile genetic elements³⁹. Fragments of *pheV*-PAI are virulence-associated and highly homologous to chromosomal regions of other entero- and uropathogenic *E. coli*⁴⁰. While DNA sequence and the virulence-associated fragments are a perfect match in *pheV*-PAIs of *E. coli* F8316/41 and *E. coli* NCTC11105 (99.97% identity), they are distinct from the remaining *E. coli* strains listed earlier. Another major distinction between *E. coli* F8316/41, *E. coli* NCTC11105 and the rest of the K15 capsular gene cluster containing *E. coli* strains is the relative position of *serA*, *pgk*, *pheV* and K15 capsular gene cluster. In *E. coli* F8316/41 and *E. coli* NCTC11105, *serA* and *pgk* genes are located within the *pheV*-PAI with the *pheV* gene placed 44,010 and 58,008 bp upstream to *pgk* and *serA* respectively, while the K15 gene cluster is 32,765 bp downstream to the *serA* gene. On the other hand, the *pheV*-PAI in the remaining six *E. coli* strains (FORC_031, FMU073332, 536, 743, MRY15-131 and MRY15-117) is located upstream to *pgk* which is in turn located upstream to the *serA* gene.

The K15 capsular gene cluster in *E. coli* F8316/41 is flanked by genes *yeeUV* (type IV toxin-antitoxin family)⁴¹ and *gspM-C* (type II secretion system)⁴². Presence of type II secretion system downstream to capsule gene cluster is the only common feature among all K15 *E. coli* strains discussed here. Pix fimbria-encoding gene cluster and the phosphoglycerate transporter system identified in the *pheV*-PAI of *E. coli* 536¹⁹ are absent from the *pheV*-PAI of *E. coli* F8316/41. Further in-depth discussion on the composition of *E. coli* F8316/41 *pheV*-PAI is not in the scope of this article. In summation of findings from the genetic analysis section, we identified the K15 capsular gene cluster from *E. coli* F8316/41 as a part of *pheV* associated pathogenicity island. Although the location of PAI in proximity to *serA* gene is consistent with observations from other *E. coli* strains, a different order of arrangement and orientation of genes in PAIs of *E. coli* F8316/41 and NCTC11105 strains point to a divergent evolution from the base *E. coli* genome. As *E. coli* F8316/41 is known to be entero- and uropathogenic, a detailed structural study of PAIs and their genomic elements can help understand the origins of virulence and their role in development of disease.

Organization of the *E. coli* K15 strain F8316/41 capsule gene cluster. The arrangement of genes in *E. coli* K15 capsule gene cluster from strain F8316/41 (Fig. 6) shares similarities with other Group 2 capsule gene clusters of the ABC transporter dependent pathway⁴³. The gene cluster is made up of three regions, with region 1 containing a total of seven ORFs (*kpsE*, *-E*, *-D*, *-U*, *-C* and two unknown putative ORFs). The first five ORFs have a high degree of sequence similarity to other region 1 genes from group 2 capsule forming *E. coli*, except for *kpsC*, which is a truncated version of *kpsC*. *KpsC* was reported to be a non-functional gene, inactivated due to truncation.

Regions 2 and 3 are transcribed as a single polycistronic mRNA in the opposite direction to that of region 1, with a JUMPstart sequence located upstream of Region 3⁴⁴. Region 3 codes for KpsM and KpsT, that are homologous to their counterparts of other group 2 capsules. Region 2 is made-up of seven ORFs and flanked by regions 1 and 3. ORFs 1 through 5 are capsule specific genes followed by *kpsC* and *kpsS* that are usually grouped with region 1 genes in group 2 capsule gene clusters. However, these two genes are described to be homologues to *kpsC* and *S* from group 3 capsule forming *E. coli* (K10), and are essential for K15 capsule production¹⁹. A low G + C content (30.9%) of regions 2 and 3 compared to region 1 (50.9%) of K15 capsular gene cluster was used as evidence to hypothesize origin of these DNA regions from different sources¹⁹. Functional analysis of capsule specific genes in region 2 was initially performed using a BLASTP search, which identified ORF 1 as a glycosyltransferase and ORF 5 as an α/β hydrolase and the remaining three ORFs as hypothetical *E. coli* proteins without any known functions. As region 2 genes are known to be capsule structure specific, using the K15 PS structure described in the earlier sections, we hypothesized the presence an α -KDOp transferase, an α -GlcNAc transferase and an *O*-acetyltransferase among the five ORFs. Although capsules with acetylation and α -GlcNAc units are common, α -KDOp is unusual with only two other capsules reported^{10,13}.

Enzymes that catalyze capsular *O*-acetylation in bacteria can be classified into one of the following two protein families⁴⁵. The first one is a hexapeptide repeat family, that contains tandem repeats of (LIV)(GAED)₂(STAV) X consensus sequence and folds into a left-handed β -helix domain (also termed L β H family)⁴⁶. Members of this family include C₅sF (both OatW and OatY)⁴⁷ from *N. meningitidis* serogroups W and Y⁴⁸, NeuO from *E. coli*

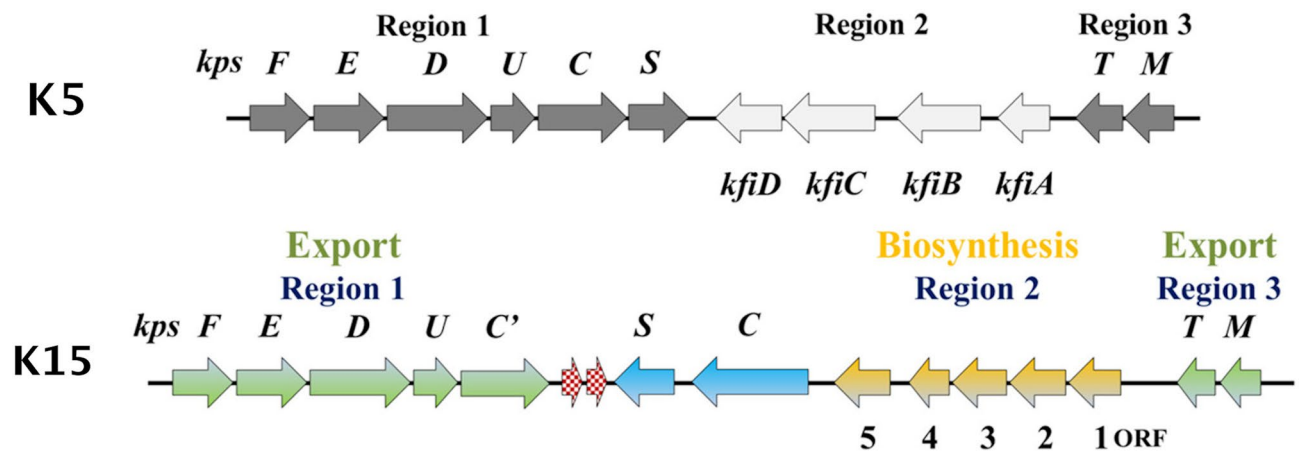


Figure 6. Genetic organization of *E. coli* K15 (F8316-14:O6:K15:H16) capsule gene cluster and its similarity to *E. coli* K5 (Group 2). Regions 1 and 3 (green) are involved in capsule transport across the inner membrane. Region 2 (yellow) is serotype specific, encoding proteins for capsule biosynthesis. ORFs in blue represent the second set of *kpsS* and *kpsC* genes that are of Group 3 origin. Red speckled ORFs are 5' and 3' regions of IS630 family transposase. A single point mutation to form a stop codon, resulted in the breakup of a single ORF into 2 fragments.

K1⁴⁹ and NeuD from *S. agalactiae*⁵⁰. In addition to the characteristic hexapeptide repeat, enzymes of this family arrange into homotrimers with an active site at each interface. The second family of capsular *O*-acetyltransferases is α/β -hydrolase family. They feature a Ser-Asp-His catalytic triad with a conformationally strained serine located in a conserved nucleophile elbow motif (GXSXGG). Previously reported members of this family are CssE (OatC) and CsaC (SacC/MynC) from *N. meningitidis* serogroups C and A respectively^{47,51}. Although crystal structure data are lacking for the α/β -hydrolase family, both CssE and CsaC have been thoroughly investigated and reported to catalyze *O*-acetylation of their respective capsular polysaccharide. Upon bioinformatics analysis (BLASTP, pfam, InterPro, etc.)⁵²⁻⁵⁴, of the five serogroup specific ORFs in region 2 of K15 capsular gene cluster, we identified ORF5 (GenBank accession number: NDK77127.1) to contain an α/β hydrolase fold (pfam-A family: UPF0227 and clan: CL0028, InterPro homologous family: IPR029058, CATH code: 3.40.50.1820) ORF5 encodes a 315 amino acid protein. Sequence alignment with CssE and CsaC (Supplementary Fig. S3) further highlighted a highly conserved catalytic triad composed of Ser-179, Asp-253 and His-281 in addition to the presence of Ser-179 in a conserved nucleophile elbow motif (GGS¹⁷⁹MGG). α/β hydrolases are a versatile family of enzymes that primarily catalyze hydrolytic reactions through a double-displacement mechanism (hydrolases, thioesterases, haloperoxidases, halogenases etc.). There is considerable structural and experimental evidence for α/β hydrolase family catalyzing *O*-acetylation reaction using a Ser-Asp-His catalytic triad (homoserine *O*-acetyltransferase from *H. influenzae*)^{55,56}. In spite of the sequence similarity between ORF5, OatC and SacC being around 30%, grouping of the three proteins into the same family via bioinformatics analysis, presence of a highly conserved catalytic triad, nucleophile elbow motif, and finally the absence of *O*-acetylation related protein structures in other region 2 ORFs is consistent with ORF5 being the capsular *O*-acetyltransferase.

Identification of specific glycosyl transferases in region 2 for transfer of GlcNAc and KDO was done by searching for capsules from other bacteria, that contained monosaccharide units and glycosidic linkages similar to the K15 PS. In this process, the CPS from *Actinobacillus pleuropneumoniae* serogroup 5a str. J45 (with 6)- α -D-GlcNAc-(1 \rightarrow 5)- β -KDOp-(2 \rightarrow repeating units) was identified containing a GlcNAc linked to the C5 of KDOp, in its α -anomeric conformation⁵⁷. In addition to this CPS structural similarity the capsular gene cluster of *A. pleuropneumoniae* shares a high degree of homology with ABC transporter dependent capsule export genes from *H. influenzae* type b, *N. meningitidis* group B, and *E. coli*⁵⁸. The capsule specific region 2 of *A. pleuropneumoniae* serogroup 5a contains four genes, *cps5A*, *5B*, *5C* and *5D* (sequence AF053723.1)⁵⁹, of which only *cps5A*, (GenBank accession number: AAC26630.1) classified as a putative glycosyltransferase exhibited a 46.2% sequence similarity with ORF1 (GenBank accession number: NDK77131.1) of K15 *E. coli*. This observation along with the absence of similarities between any other region 2 genes of either organisms suggests that ORF1 might encode for an α -GlcNAc transferase. Bioinformatic analysis using primary protein sequence predicted ORF1 to be a member of GT4 family of glycosyltransferases (pfam-A family: Glyco trans_1_4 and clan: CL0113, CDD: cd03801-GT4 PimA-like, CATH code: 3.40.50.2000). As supporting evidence to this hypothesis, a characteristic signature motif, typical to retaining α -glycosyltransferases was observed in the C-terminal end of ORF1 (S211PYEGGPA-CLPEALA225). This conserved secondary structure is an EX₇E motif, first reported using hydrophobic cluster analysis⁶⁰⁻⁶². The EX₇E motif in ORF1 of *E. coli* K15 perfectly aligns with S224SHEGGPANIP₇EALA238 in Cps5A of *A. pleuropneumoniae* (Supplementary Fig. S4). Several other retaining glycosyltransferases belonging to GT4 families carry this conserved motif (for example: *E. coli* lipopolysaccharide α -glucosyltransferase WaaG; *Mycobacterium smegmatis* α -mannosyltransferase, PimA and *Corynebacterium glutamicum* glycosyltransferase, MshA) (Supplementary Fig. S4).

The K15 structure reported previously contains a β -KDO in the repeat unit, thus predicting the presence of β -KDO transferase. Our analysis of the gene cluster did not find evidence of a sequence belonging to the β -KDO transferase family GT-99. The uncommon presence of α -KDOp in the capsular polysaccharide made it difficult to use the same approach we employed in identifying α -GlcPNAc transferase.

Bioinformatic analysis using CDD, Pfam, dbCAN and InterPro did not generate any hits for ORFs 2,3, and 4. As a result, the second capsule specific glycosyl transferase (α -KDOp transferase) was identified using the Phyre² web portal for protein structure modeling and functional analysis. This tool generated hits for ORF4 (GenBank accession number: NDK77128.1) with 100 and 99.5% confidence to Maf (Motility associated factor) glycosyltransferase, from *Magnetospirillum magneticum* AMB-1, and α -2,3/8-sialyltransferase CstII from *Campylobacter jejuni*, respectively. Though the sequence identity was only 19%, models generated with high confidence have been shown to be predictive of structural features and function⁶³. Maf glycosyltransferase belongs to a class of bacterial glycosyltransferases involved in transfer of nonulosonic acids like moieties to an acceptor⁶⁴. The Maf central domain has been shown to exhibit similarity to a sialyltransferase for *C. jejuni*⁶⁴. Taking all these findings into account, we predict ORF4 to be an α -KDOp transferase with no structural homology to currently documented glycosyltransferases. In summary, we have identified genes in the K15 gene cluster that potentially encode an *O*-acetyltransferase, an *N*-acetylglucosamine transferase, and a KDO transferase consistent with the structure we report. The β -KDO transferases encoded by CPS gene clusters have been reported to belong to the glycosyltransferase family GT99 and have a different fold than KDO transferases associated with LPS biosynthesis¹⁴. The relationship of the α -KDO transferases of *E. coli* K6, K15, and K16 to these two different families of glycosyltransferases awaits further structural and function analysis.

Methods

Growth of the bacteria and preparation of polysaccharide. *Escherichia coli* strain F8316/41 (O6:K15:H16) was obtained from Drs. F. and I. Orskov at Staten Serum Institute, Copenhagen, Denmark. For the isolation of polysaccharide, bacteria were grown on a low molecular weight medium and the polysaccharide was purified by published methods⁶⁵. During isolation, volumes were kept to a minimum to improve yields. Carboxylate reduced polysaccharide was prepared by the carbodiimide-borohydride method of Taylor⁶⁶.

NMR spectroscopy. 20 mg of K15 polysaccharide were dissolved in 0.6 mL of 99.9% D₂O. The resulting solution was approximately neutral. De-*O*-acetylation of K15 was achieved by increasing the solution pH to ~12 by addition of NaOH from a stock solution. NMR experiments were run on a Bruker 700 MHz NMR instrument equipped with a triple gradient TCI cryoprobe. The temperature was set to 50 °C to reduce the line widths. The internal reference standard was ~0.1 (wt %) DSS-d₆. NMR spectra were obtained using Bruker's TopSpin 3.5 software (www.bruker.com), analysis was performed with both Topspin and Mnova 11 (www.mestrelab.com). A spin-lock time of 60 ms was used in the HSQC-TOCSY experiment. For the evolution of long-range couplings in the ¹H,¹³C HMBC experiment a *J*_{H,C} coupling constant of 6 Hz was used.

Other analytical methods. Sugar components were detected by paper chromatography, paper electrophoresis, gas chromatography and automated sugar analysis. Thin layer chromatography was performed on TLC-cellulose plates in butanol-pyridine-water, 6/4/3, v/v/v⁶⁷. Hexosamine was released by hydrolysis in 4 N HCl. KDO was quantified in hydrolysates by the thiobarbituric acid assay. Alditol acetates were prepared and analyzed as described previously⁶⁵.

Glycosyl composition. Glycosyl composition analysis was performed by combined gas chromatography/mass spectrometry (GC/MS) of the per-*O*-trimethylsilyl (TMS) derivatives of the monosaccharide methyl glycosides produced from the sample by acidic methanolysis as described previously by Santander et al.⁶⁸. Briefly, the samples (230 and 240 μ g) were heated with methanolic HCl in a sealed screw-top glass test tube for 17 h at 80 °C. After cooling and removal of the solvent under a stream of nitrogen, the samples were treated with a mixture of methanol, pyridine, and acetic anhydride for 30 min to re-*N*-acetylate the hexosamines. The solvents were evaporated, and the samples were derivatized with Tri-Sil (Pierce) at 80 °C for 30 min. GC/MS analysis of the TMS methyl glycosides was performed on an Agilent 7890A GC interfaced to a 5975C MSD, using a Supelco Equity-1 fused silica capillary column (30 m 0.25 mm ID).

Degradation procedures. Oligosaccharides were prepared by mild acid hydrolysis in 1% acetic acid at 100 °C for 1 h. The resulting disaccharide was purified by gel filtration on Bio-Gel P-2 in 0.1 M ammonium acetate and subsequent paper electrophoresis in pyridine acetate, pH 5.4. K15 was oxidized with excess sodium metaperiodate for 40 h at 4 °C and desalted after quenching. Oxidized polysaccharide was reduced with sodium borohydride at pH 7.0 on a pH stat. Beta elimination of periodate oxidized polysaccharide was performed as follows. Oxidized polysaccharide was dialyzed, lyophilized, and then treated with 0.1 N sodium hydroxide at 37 °C for 2 h. The reaction mixture was applied to a TLC cellulose plate in butanol-pyridine-water, 6/4/3, v/v/v.

Hexosamine released by acid hydrolysis was degraded to a pentose by ninhydrin degradation in pyridine as described previously²⁰. The products were identified by automated sugar analysis.

Glycosyl linkage analysis. For glycosyl linkage analysis, the samples were permethylated, reduced, hydrolyzed under mild conditions, reduced, hydrolyzed again and acetylated; and the resultant partially methylated alditol acetates (PMAAs) analyzed by gas chromatography-mass spectrometry (GC-MS). The procedure is a slight modification of the one described by Willis et al.⁶⁹.

About 500 µg of the samples were used for linkage analysis. The samples were suspended in 200 µl of dimethyl sulfoxide and left to stir for 1 day. Permethylation of the sample was affected by two rounds of treatment with sodium hydroxide (15 min) and methyl iodide (45 min). The permethylated sample carboxylic acids were reduced by adding 200 µl of a 5 mg/ml solution of LiBD₄ in 90% THF and reacting overnight at room temperature, followed by 1 h at 100 °C. The samples were then hydrolyzed using 0.1 M TFA (0.5 h in sealed tube at 100 °C), reduced with NaBD₄, hydrolyzed again using 2 M TFA (2 h in sealed tube at 100 °C) and acetylated using acetic anhydride/TFA. The resulting PMAAs were analyzed on an Agilent 7890A GC interfaced to a 5975C MSD (mass selective detector, electron impact ionization mode); separation was performed on a 30 m Supelco SP-2331 bonded phase fused silica capillary column.

Whole genome sequencing and capsule cluster analysis. The chromosomal DNA isolated from *E. coli* str. F8316/41 was sequenced using single-molecule real-time (SMRT) sequencing technology³⁴. The raw sequencing data was processed using Canu v. 1.5⁷⁰. Canu is a long read single-molecule sequence de novo assembly tool for PacBio and/or Nanopore reads. This process resulted in five contiguous DNA segments with base lengths as follows: 2,558,639, 2,393,415, 99,428, 88,757, and 29,847 bases (GenBank accession numbers: JAADZB010000001 – JAADZB010000005). The second contiguous DNA segment with 2,393,415 bases was amplified in 20 different regions using forward and reverse primers (see Supplementary Information) to perform Sanger sequencing. The contiguous DNA sequences were annotated using NCBI's Prokaryotic Genome Annotation Pipeline (PGAP) and submitted to GenBank⁷¹. The K15 capsular gene cluster and the *pheV*-PAI genetic region were analyzed and compared with genomes of other *E. coli* K15 strains using BLASTN³⁸ and BLASTP⁵³. The ORFs in the K15 gene cluster were generated using ORFfinder (<https://www.ncbi.nlm.nih.gov/orffinder/>). A wide array of bioinformatic resources like the pfam database⁵², CDD search⁷², InterPro classification of protein families⁵⁴, CATH/Gene3D v4.2⁷³, Phyre² protein fold recognition server⁷⁴ and dbCAN meta server⁷⁵ for automated CAZyme annotation were used for identification and functional characterization of capsule specific region 2 genes.

Received: 19 February 2020; Accepted: 13 July 2020

Published online: 28 July 2020

References

1. Holst, O., Moran, A.P. & Brennan, P.J. in *Microbial Glycobiology: Structures, Relevance and Applications*. (ed. A.P. Moran) 3–13 (Academic Press, London; 2009).
2. Jann, K. & Jann, B. The K antigens of *Escherichia coli*. *Prog. Allergy* **33**, 53–79 (1983).
3. Willis, L. M. & Whitfield, C. Structure, biosynthesis, and function of bacterial capsular polysaccharides synthesized by ABC transporter-dependent pathways. *Carbohydr. Res.* **378**, 35–44 (2013).
4. Lindberg, B. Components of bacterial polysaccharides. *Adv. Carbohydr. Chem. Biochem.* **48**, 279–318 (1990).
5. Jann, B., Hofmann, P. & Jann, K. Structure of the 3-deoxy-D-manno-octulosonic acid-(KDO)-containing capsular polysaccharide (K14 antigen) from *Escherichia coli* 06:K14:H31. *Carbohydr. Res.* **120**, 131–141 (1983).
6. Jann, B. & Jann, K. Structure and biosynthesis of the capsular antigens of *Escherichia coli*. *Curr. Top. Microbiol. Immunol.* **150**, 19–42 (1990).
7. Jann, K. & Jann, B. Polysaccharide antigens of *Escherichia coli*. *Rev. Infect. Dis.* **9**, S517–S526 (1987).
8. Schmidt, M. A. & Jann, K. Structure of the 2-keto-3-deoxy-D-manno-octonic-acid-containing capsular polysaccharide (K12 antigen) of the urinary-tract-infective *Escherichia coli* O4:K12:H. *Eur. J. Biochem.* **131**, 509–517 (1983).
9. Griffiths, A. J. & Davies, D. B. Type-specific carbohydrate antigens of pathogenic bacteria. Part 1: enterobacteriaceae. *Carbohydr. Polym.* **14**, 241–279 (1990).
10. Jennings, H. J., Rosell, K. G. & Johnson, K. G. Structure of the 3-deoxy-D-manno-octulosonic acid-containing polysaccharide (K6 antigen) from *Escherichia coli* LP 1092. *Carbohydr. Res.* **105**, 45–56 (1982).
11. Birnbaum, G. I., Roy, R., Brisson, J. R. & Jennings, H. J. Conformations of ammonium 3-deoxy-D-manno-2-octulosonate (KDO) and methyl α- and β-ketopyranosides of KDO: x-ray structure and proton NMR analyses. *J. Carbohydr. Chem.* **6**, 17–39 (1987).
12. Neszmelyi, A., Jann, K., Messner, P. & Unger, F. Constitutional and configurational assignments by carbon-13 NMR spectroscopy of *Escherichia coli* capsular polysaccharides containing ribose and 3-deoxy-D-manno-2-octulosonic acid (KDO). *J. Chem. Soc. Chem. Commun.*, 1017–1019 (1982).
13. Lenter, M., Jann, B. & Jann, K. Structure of the K16 antigen from *Escherichia coli* O7:K16:H-, a Kdo-containing capsular polysaccharide. *Carbohydr. Res.* **197**, 197–204 (1990).
14. Ovchinnikova, O. G. *et al.* Bacterial beta-Kdo glycosyltransferases represent a new glycosyltransferase family (GT99). *Proc. Natl. Acad. Sci. USA* **113**, E3120–3129 (2016).
15. Binsztein, N. *et al.* Colonization factors of enterotoxigenic *Escherichia coli* isolated from children with diarrhea in Argentina. *J. Clin. Microbiol.* **29**, 1893–1898 (1991).
16. Blum, G., Marre, R. & Hacker, J. Properties of *Escherichia coli* strains of serotype O6. *Infection* **23**, 234–236 (1995).
17. Orskov, F. *et al.* Special *Escherichia coli* serotypes among enterotoxigenic strains from diarrhoea in adults and children. *Med. Microbiol. Immunol.* **162**, 73–80 (1976).
18. Orskov, I. & Orskov, F. Special O:K: H serotypes among enterotoxigenic *E. coli* strains from diarrhea in adults and children. Occurrence of the CF (colonization factor) antigen and of hemagglutinating abilities. *Med. Microbiol. Immunol.* **163**, 99–110 (1977).
19. Schneider, G. *et al.* The pathogenicity island-associated K15 capsule determinant exhibits a novel genetic structure and correlates with virulence in uropathogenic *Escherichia coli* strain 536. *Infect. Immun.* **72**, 5993–6001 (2004).
20. Stoffyn, P. J. & Jeanloz, R. W. Identification of amino sugars by paper chromatography. *Arch. Biochem. Biophys.* **52**, 373–379 (1954).
21. Jennings, H. J. & Smith, I. C. P. 3 Polysaccharide structures using carbon-13 nuclear magnetic resonance. *Methods Enzymol.* **50**, 39–50 (1978).
22. Lerner, L. & Bax, A. Sensitivity-enhanced two-dimensional heteronuclear relayed coherence transfer nmr-spectroscopy. *J. Magn. Reson.* **69**, 375–380 (1986).
23. Lerner, L. & Bax, A. Application of new, high-sensitivity 1H–13C-N.M.R.- spectral techniques to the study of oligosaccharides. *Carbohydr. Res.* **166**, 35–46 (1987).

24. Bhattacharjee, A. K., Jennings, H. J. & Kenny, C. P. Structural elucidation of the 3-deoxy-D-manno-octulosonic acid containing meningococcal 29-e capsular polysaccharide antigen using carbon-13 nuclear magnetic resonance. *Biochemistry* **17**, 645–651 (1978).
25. Jennings, H. J., Rosell, K. G. & Johnson, K. G. Structure of the 3-deoxy-D-manno-octulosonic acid-containing polysaccharide (K6 antigen) from *Escherichia coli* LP 1092. *Carbohydr Res* **105**, 45–56 (1982).
26. Kogler, H. O. W. Sorensen, G. Bodenhausen, and RR Ernst. *J. Magn. Reson.* **55**, 1 (1983).
27. Bax, A. & Summers, M. F. Proton and carbon-13 assignments from sensitivity-enhanced detection of heteronuclear multiple-bond connectivity by 2D multiple quantum NMR. *J. Am. Chem. Soc.* **108**, 2093–2094 (1986).
28. Birnbaum, G. I., Roy, R., Brisson, J.-R. & Jennings, H. J. Conformations of ammonium 3-Deoxy-D-manno-2-octulosonate (KDO) and methyl α - and β -ketopyranosides of KDO: x-ray structure and 1H NMR analyses. *J. Carbohydr. Chem.* **6**, 17–39 (1987).
29. Kohlbrenner, W. E. & Fesik, S. W. Determination of the anomeric specificity of the *Escherichia coli* CTP: CMP-3-deoxy-D-manno-octulosonate cytidyltransferase by 13C NMR spectroscopy. *J. Biol. Chem.* **260**, 14695–14700 (1985).
30. Bush, C. A. High resolution NMR in the determination of structure in complex carbohydrates. *Bull. Magn. Reson.* **10**, 73–95 (1988).
31. Pomin, V. H. *Unravelling Glycobiology by NMR Spectroscopy* 63–98 (Intech, London, 2012).
32. Yu, B. E. A. More accurate $^1J_{CH}$ coupling measurement in the presence of $^3J_{HH}$ strong coupling in natural abundance. *J. Magn. Reson.* **215**, 10–22 (2012).
33. Shashkov, A. S., Arbatsky, N. P. & Knirel, Y. A. Structures and genetics of Kdo-containing O-antigens of *Cronobacter sakazakii* G2706 and G2704, the reference strains of serotypes O5 and O6. *Carbohydr. Res.* **346**, 1924–1929 (2011).
34. Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138 (2009).
35. Oerskov, I., Sharma, V. & Oerskov, F. Genetic mapping of the K1 and K4 antigens (L) of *Escherichia coli*. Non-allelism of K(L) antigens with K antigens of O8:K27(A), O8:K8(L) and O9:K57(B). *Acta Pathol. Microbiol. Scand. Sect. B Microbiol.* **84B**, 125–131 (1976).
36. Orskov, I. & Nyman, K. Genetic mapping of the antigenic determinants of two polysaccharide K antigens, K10 and K54, in *Escherichia coli*. *J. Bacteriol.* **120**, 43–51 (1974).
37. Vimr, E. R. Map position and genomic organization of the kps cluster for polysialic acid synthesis in *Escherichia coli* K1. *J. Bacteriol.* **173**, 1335–1338 (1991).
38. Camacho, C. *et al.* Blast+: architecture and applications. *BMC Bioinf.* **10**, 421 (2009).
39. Dobrindt, U. *et al.* Analysis of genome plasticity in pathogenic and commensal *Escherichia coli* isolates by use of DNA arrays. *J. Bacteriol.* **185**, 1831–1840 (2003).
40. Beatson, S. A. *et al.* Molecular analysis of asymptomatic bacteriuria *Escherichia coli* strain VR50 reveals adaptation to the urinary tract by gene acquisition. *Infect. Immun.* **83**, 1749–1764 (2015).
41. Brown, J. M. & Shaw, K. J. A novel family of *Escherichia coli* toxin-antitoxin gene pairs. *J. Bacteriol.* **185**, 6600–6608 (2003).
42. Patrick, M., Gray, M. D., Sandkvist, M. & Johnson, T. L. Type II secretion in *Escherichia coli*. *EcoSal Plus* **4**, 1–18 (2014).
43. Roberts, I. S. The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annu. Rev. Microbiol.* **50**, 285–315 (1996).
44. Bailey, M. J. A., Hughes, C. & Koronakis, V. RfaH and the ops element, components of a novel system controlling bacterial transcription elongation. *Mol. Microbiol.* **26**, 845–851 (1997).
45. Muehlenhoff, M. & Bergfeld, A. K. 209–235 (Bentham Science Publishers Ltd., 2014).
46. Raetz, C. R. H. & Roderick, S. L. A left-handed parallel β helix in the structure of UDP-N-acetylglucosamine acyltransferase. *Science (Washington, D. C.)* **270**, 997–1000 (1995).
47. Bergfeld, A. K. *et al.* The polysialic acid-specific O-acetyltransferase OatC from *Neisseria meningitidis* serogroup C evolved apart from other bacterial sialate O-acetyltransferases. *J. Biol. Chem.* **284**, 6–16 (2009).
48. Lee, H. J. *et al.* Structural and kinetic characterizations of the polysialic acid O-acetyltransferase OatWY from *Neisseria meningitidis*. *J. Biol. Chem.* **284**, 24501–24511 (2009).
49. Schulz, E. C., Bergfeld, A. K., Ficner, R. & Muehlenhoff, M. Crystal structure analysis of the polysialic acid specific O-acetyltransferase NeuO. *PLoS ONE* **6**, e17403 (2011).
50. Lewis, A. L., Hensler, M. E., Varki, A. & Nizet, V. The group B streptococcal sialic acid O-acetyltransferase is encoded by neuD, a conserved component of bacterial sialic acid biosynthetic gene clusters. *J. Biol. Chem.* **281**, 11186–11192 (2006).
51. Gudlavalleti, S. K. *et al.* The *Neisseria meningitidis* serogroup A capsular polysaccharide O-3 and O-4 acetyltransferase. *J. Biol. Chem.* **279**, 42765–42773 (2004).
52. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
53. Gish, W. & States, D. J. Identification of protein coding regions by database similarity search. *Nat. Genet.* **3**, 266–272 (1993).
54. Mitchell, A. L. *et al.* InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucl. Acids Res.* **47**, D351–D360 (2019).
55. Mirza, I. A., Nazi, I., Korczynska, M., Wright, G. D. & Berghuis, A. M. Crystal structure of homoserine transacetylase from *Haemophilus influenzae* reveals a new family of α/β -hydrolases. *Biochemistry* **44**, 15768–15773 (2005).
56. Wang, M. *et al.* Crystal structure of homoserine O-acetyltransferase from *Leptospira interrogans*. *Biochem. Biophys. Res. Commun.* **363**, 1050–1056 (2007).
57. Altman, E., Brisson, J. R. & Perry, M. B. Structure of the capsular polysaccharide of *Haemophilus pleuropneumoniae* serotype 5. *Eur. J. Biochem.* **170**, 185–192 (1987).
58. Ward, C. K. & Inzana, T. J. Identification and characterization of a DNA region involved in the export of capsular polysaccharide by *Actinobacillus pleuropneumoniae* serotype 5a. *Infect. Immun.* **65**, 2491–2496 (1997).
59. Ward, C. K., Lawrence, M. L., Veit, H. P. & Inzana, T. J. Cloning and mutagenesis of a serotype-specific DNA region involved in encapsulation and virulence of *Actinobacillus pleuropneumoniae* serotype 5a: concomitant expression of serotype 5a and 1 capsular polysaccharides in recombinant *A. pleuropneumoniae* serotype 1. *Infect. Immun.* **66**, 3326–3336 (1998).
60. Amor, P. A. & Whitfield, C. Molecular and functional analysis of genes required for expression of group IB K antigens in *Escherichia coli*: characterization of the his-region containing gene clusters for multiple cell-surface polysaccharides. *Mol. Microbiol.* **26**, 145–161 (1997).
61. Geremia, R. A., Petroni, E. A., Ielpi, L. & Henrissat, B. Towards a classification of glycosyltransferases based on amino acid sequence similarities: prokaryotic α -mannosyltransferases. *Biochem. J.* **318**, 133–138 (1996).
62. Saxena, I. M., Brown, R. M. Jr., Fevre, M., Geremia, R. A. & Henrissat, B. Multidomain architecture of β -glycosyl transferases: implications for mechanism of action. *J. Bacteriol.* **177**, 1419–1424 (1995).
63. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
64. Sulzenbacher, G. *et al.* Glycosylate and move! The glycosyltransferase Maf is involved in bacterial flagella formation. *Environ. Microbiol.* **20**, 228–240 (2018).
65. Vann, W. F. & Jann, K. Structure and serological specificity of the K13-antigenic polysaccharide (K13 antigen) of urinary tract-infective *Escherichia coli*. *Infect. Immun.* **25**, 85–92 (1979).
66. Taylor, R. L. & Conrad, H. E. Stoichiometric depolymerization of polyuronides and glycosaminoglycuronans to monosaccharides following reduction of their carbodiimide-activated carboxyl groups. *Biochemistry* **11**, 1383–1388 (1972).

67. Vann, W. F. *et al.* Serological, chemical, and structural analyses of the *Escherichia coli* cross-reactive capsular polysaccharides K13, K20, and K23. *Infect. Immun.* **39**, 623–629 (1983).
68. Santander, J. *et al.* Mechanisms of intrinsic resistance to antimicrobial peptides of *Edwardsiella ictaluri* and its influence on fish gut inflammation and virulence. *Microbiology* **159**, 1471–1486 (2013).
69. Willis, L. M. *et al.* Conserved glycolipid termini in capsular polysaccharides synthesized by ATP-binding cassette transporter-dependent pathways in gram-negative pathogens. *Proc. Natl. Acad. Sci. USA* **110**, 7868–7873 (2013).
70. Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
71. Tatusova, T. *et al.* NCBI prokaryotic genome annotation pipeline. *Nucl. Acids Res.* **44**, 6614–6624 (2016).
72. Marchler-Bauer, A. *et al.* CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucl. Acids Res.* **45**, D200–D203 (2017).
73. Lewis, T. E. *et al.* Gene3D: Extensive prediction of globular domains in proteins. *Nucl. Acids Res.* **46**, D435–D439 (2018).
74. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
75. Zhang, H. *et al.* DbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucl. Acids Res.* **46**, W95–W101 (2018).

Acknowledgements

Permethylated and composition analysis was performed by the Analytical Group at CCRC at the University of Georgia. The work performed at CCRC was supported by the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, U.S. Department of Energy grant (DE-SC0015662) to Parastoo Azadi at the Complex Carbohydrate Research Center. The PacBio single-molecule real-time (SMRT) sequencing was performed at Biomedical Research Core Facilities DNA sequencing core, University of Michigan and the raw data was processed using Canu v. 1.5 with the help of Luis Santana-Quintero and Vahan Simonyan at High-performance Integrated Virtual Environment (HIVE) Laboratory, Center for Biologics Evaluation and Research, U.S. FDA.

Author contributions

H.F.A. and S. F. performed the NMR analysis, V.V. did the genetic characterization, S.F. V.V. and W.V. were responsible for performing and procuring the chemical analysis, F.L. purified K15 polysaccharide, H.F.A., V.V., D.I.F., and W.F.V. prepared the manuscript. H.F.A. and V.V. made equivalent contributions to this manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-69476-z>.

Correspondence and requests for materials should be addressed to W.F.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2020