This is the published version of a paper published in *Glycobiology*.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-116533

OXFORD

Original Article

Development of the ECODAB into a relational database for *Escherichia coli* O-antigens and other bacterial polysaccharides[†]

Miguel A Rojas-Macias[2,‡], Jonas Ståhle[3,‡], Thomas Lütteke[2], and Göran Widmalm[1,3]

[2]Institute of Veterinary Physiology and Biochemistry, Justus-Liebig-University Giessen, Frankfurter Str. 100, Giessen 35392, Germany, and [3]Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, S-106 91 Stockholm, Sweden

[1]To whom correspondence should be addressed: Tel: +46-8-16-37-42; Fax: +46-8-15-49-08; e-mail: gw@organ.su.se

[†]Presented in part at the International Carbohydrate Symposium, Madrid, Spain 22–27 July 2012, P236. Development of ECODAB into a relational database.
[‡]These authors contributed equally to this work.

Abstract

Escherichia coli O-antigen database (ECODAB) is a web-based application to support the collection of *E. coli* O-antigen structures, polymerase and flippase amino acid sequences, NMR chemical shift data of O-antigens as well as information on glycosyltransferases (GTs) involved in the assembly of O-antigen polysaccharides. The database content has been compiled from scientific literature. Furthermore, the system has evolved from being a repository to one that can be used for generating novel data on its own. GT specificity is suggested through sequence comparison with GTs whose function is known. The migration of ECODAB to a relational database has allowed the automation of all processes to update, retrieve and present information, thereby, endowing the system with greater flexibility and improved overall performance. ECODAB is freely available at http://www.casper.organ.su.se/ECODAB/. Currently, data on 169 *E. coli* unique O-antigen entries and 338 GTs is covered. Moreover, the scope of the database has been extended so that polysaccharide structure and related information from other bacteria subsequently can be added, for example, from *Streptococcus pneumoniae*.

Key words: database, ECODAB, *Escherichia coli*, glycosyltransferase specificity, O-antigen

Introduction

The flagellated rod-shaped Gram-negative bacterium *Escherichia coli* colonizes the human gastrointestinal tract just a few hours after birth, where it becomes an important part of the lower gut flora. Normally innocuous, *E. coli* accompanies its host for life in a mutually beneficial relationship (Kaper et al. 2004). Unfortunately, this placid commensal association can be broken when *E. coli* acquires genetic elements encoding for virulence factors and turns pathogenic (Stenutz et al. 2006). Thereupon, formerly harmless *E. coli* strains may become responsible

for a variety of diarrheal and extra-intestinal infections in humans and animals, some of which have the potential to be fatal. In certain cases, when the immune system of the host turns ineffective or gastrointestinal barriers are violated, even non-pathogenic *E. coli* can provoke infection (Nataro and Kaper 1998; Stenutz et al. 2006).

The basis for the serological identification and classification of *E. coli* strains were set by Kauffman in 1944, based on the main antigens that the bacteria express: O (somatic lipopolysaccharide), H (flagellar) and K (capsular) (Ørskov et al. 1977; Ewing 1986). O-Antigens

341

determine O-serogroups, which have been denoted from O1 to O187 (Stenutz et al. 2006; F Scheutz, 2014, personal communication). In this typing scheme, however, seven serogroups have been removed and certain others have been subdivided, for example, into O5ab and O5ac (Urbina et al. 2005). The combination of O- and H-antigens define a serotype. Only a limited number of laboratories are able to type K-antigens; therefore, the serotyping of O and H is considered the gold standard for subdifferentiation of *E. coli* (Ørskov and Ørskov 1992; Prager et al. 2003). To date, >60 H- and 80 K-antigens have been proposed (Stenutz et al. 2006). O-Antigens coat Gram-negative bacterial cell walls as the outermost exposed domain of polysaccharides (Figure 1), where they serve as receptors for bacteriophages, are involved in immune response, provide various defense mechanisms for the bacteria (Samuel and Reeves 2003; Li et al. 2010) and identify strain-specific surface antigens, some of which are associated with high virulence (DebRoy et al. 2011).

These O-antigens are assembled by a repeated sequence of oligosaccharides called O-units, each consisting of two up to seven sugar residues. In *E. coli*, O-antigens typically range from 10 to ~25 O-units, as determined by NMR spectroscopy or MALDI-TOF mass spectrometry (Linnerborg et al. 1999a,b; Lycknert and Widmalm 2004). Four separate biosynthetic pathways have been determined for the assembly of O-antigens, namely (i) the Wzy/Wzx-, (ii) the ABC-transporter-, (iii) the synthase- and (iv) the Wzk-dependent pathway (Hug et al. 2010), of which only the first two have been found in *E. coli* (Valvano et al. 2011) with the Wzy/Wzx-dependent pathway being the most abundant. The Wzy/Wzx-dependent pathway is typically used for heteropolymers while homopolymers and some disaccharide heteropolymers are formed through the ABC-transporter-dependent pathway (Valvano et al. 2011). Three classes of proteins are involved in the biosynthesis of the O-antigen: (i) proteins responsible for the biosynthesis of sugar units, (ii) glycosyltransferases (GTs) and (iii) O-antigen processing proteins such as Wzy (polymerase), Wzx (flippase) and Wzz (chain length regulating protein); the latter confers the lipopolysaccharide modal lengths of four to >100 repeating units (Kalynych et al. 2011). The genes encoding these proteins, collectively referred to as the O-antigen gene cluster, are typically located between the *galF* and the *gnd* operon with some exceptions, most commonly

for ABC-transporter-dependent systems, where they are instead located between the *gnd* and the *his* operon. Genes encoding the biosynthesis of common sugars are found outside of the O-antigen gene cluster and usually only the presence of non-common sugars in the O-antigen structure can be predicted by analysis of the genes encoding the sugar residue biosynthesis. The process of adding O-acetyl groups and glucosylation (Wang et al. 2007) is carried out after polymerization of the O-antigen polysaccharide (O-PS) and as such the presence of O-acetyl groups does not impact the GT function specificity. The genes encoding the O-acetyl transferases are not restricted to the O-antigen gene cluster (Allison and Verma 2000; Guo et al. 2005).

The GTs constitute a very diverse family of enzymes responsible for catalyzing glycosidic bond formation by sequentially recognizing and transferring a sugar residue from an activated sugar donor to the respective acceptor (Coutinho et al. 2009). The repertoire of glycan residues, the order in which they are assembled and the type of glycosidic linkages built between them account for the great diversity of O-antigens. To accommodate for all the possible linkages, GTs show a remarkable specificity, which in turn is reflected in their peptide sequences. The precise determination of GTs' specificity is complicated by their vast variation. Moreover, biochemical determination is constrained by the limited availability of activated sugar precursors and acceptors. GT function can also be characterized by the mutation of genes to reveal the accumulated intermediate, but two problems are associated with this method: (i) the accumulated intermediate can damage the cell and (ii) the environment can influence O-antigen structure, as implied by the ability of GTs to carry out reactions with other substrates than their usual (Stevenson et al. 2008). For *E. coli*, a few cases have been reported where there is a discrepancy between the number of glycosidic linkages per repeating unit and the number of GTs. This can possibly be explained by the previously mentioned post-polymerization glucosylation (serogroups O4, O13, O73, inter alia), in some cases enzyme bifunctionality where one GT can catalyze formation of more than one linkage (serogroups O2, O139, O150, inter alia), or for some GTs a lack of an apparent function (Rocchetta et al. 1998; Samuel and Reeves 2003; Wang et al. 2005; Lundborg et al. 2010).
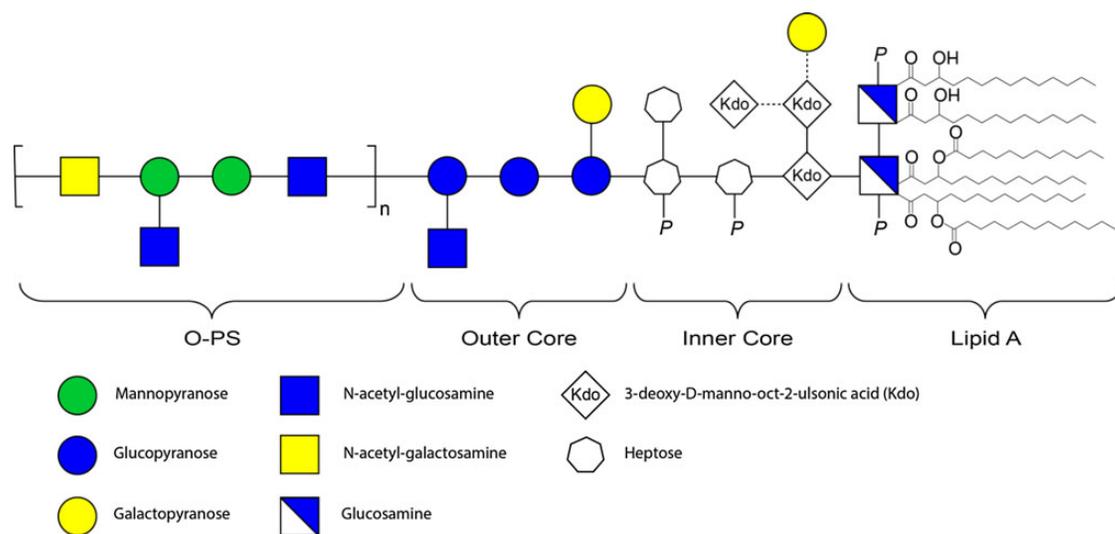

Fig 1. Schematic structure of an LPS with sugar residues in CFG-format. The O-antigen corresponds to *E. coli* O6 and the core to R2. The type of sugar residues, the order and the linkages endow O-antigens with a huge diversity.

Bioinformatics tools can be used as a means to overcome the problems with experimental methods. The *E. coli* O-antigen database (ECODAB) was created as a tool to emphasize the particularities of known *E. coli* serogroups and to facilitate the identification of common epitopes (Stenutz et al. 2006). Data on GTs were later included (Lundborg et al. 2010) and protein–protein BLAST searches (Altschul et al. 1990) were implemented with the objective of finding similarities at an amino acid sequence level between GTs with known and unknown functions. In this way, a suggestion on the probable function of GTs was possible, based on amino acid sequence data. Therefore, ECODAB has evolved from simply storing collected data to a database that can be used to produce data on its own. In the present article, a new version of ECODAB is introduced where this process has been completely automated. The sequence comparisons are also no longer restricted to those from *E. coli*, but could also include sequences from other bacteria. Moreover, the data previously stored in a file-based repository have been migrated into a relational database in order to optimize the overall system functionality and enable the possibility of future developments.

In a relational database, the major entities in a system are modelled to a set of attributes that can be visually depicted as the columns in a table. Every row in the table (a tuple or group of attribute values) is unequivocally identified by a primary key (Figure 2A), an attribute whose values are unique in the table. Relationships are established by copying the primary key of one entry to a related entry in another table. The primary key in the second table is referred to as a foreign key. When retrieving data, the common values shared by primary and foreign keys are used to combine the columns of different tables in order to reach any value in the database (Figure 2B). However, before being displayed, the data retrieved from the database need to be ordered and processed to generate useful information.

Results and discussion

In ECODAB, all relevant data regarding serogroups are displayed in only one interface, which is organized in segments where each one describes different aspects about a serogroup: O-antigen description, GTs (Figure 3A), NMR, enzymes, non-sugar components and literature (Figure 3B).

As its name indicates, the O-antigen description section lists the attributes that identify *E. coli* strains, such as H- and K-antigens and the sequence, linkage and branching of O-antigen structures. The GT section condenses all the data on the GTs that have been determined for the corresponding serogroup and provides links to other databases such as NCBI (Benson et al. 2011) and CAZy (Lombard et al. 2014). To facilitate the interpretation of results, a color code is used as follows: (i) green when the function has been experimentally
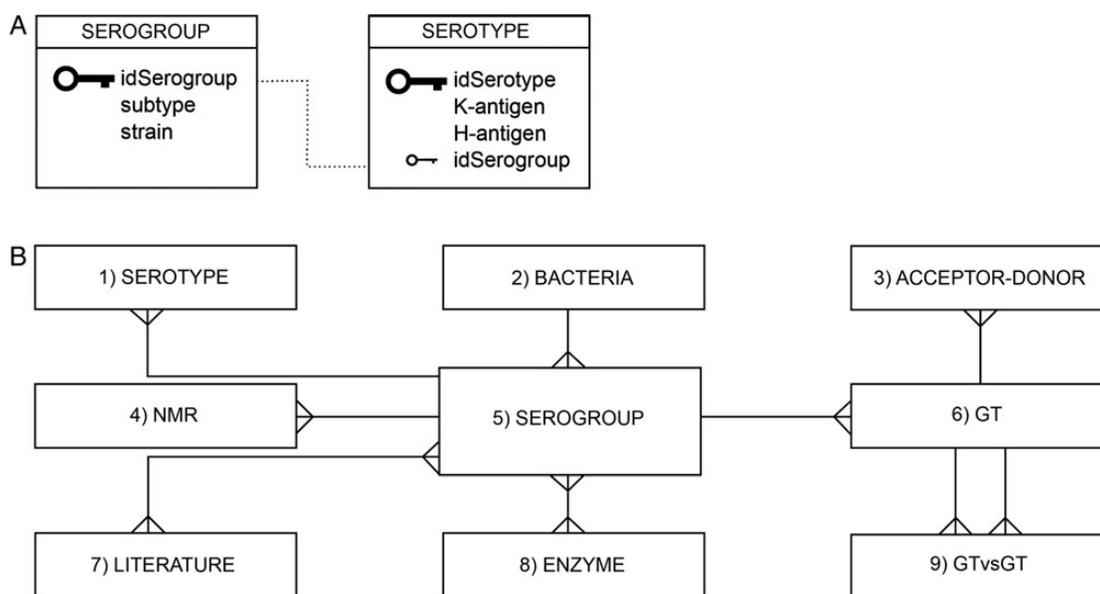
Fig. 2. (A) In a relational database, the data are stored in tables. Every table must include at least one column whose values are unique, a primary key, so that every entry in the table can be unequivocally identified. Primary keys, shown as a larger key, are copied to related entries in other tables, where they are referred to as foreign keys. Relationships between tables are in this way established. The retrieval of data stored in different tables is possible through the combination of entries based on the common keys they share. **(B)** The blocks represent the main tables in the database and the lines indicate the relationships between them.

Fig. 3. Example of the interface of a serogroup in ECODAB: (**A**) O-antigen structure, general data about the serogroup and brief description of the determined GTs for the respective serogroup. Colors are used to indicate the status of a GT: green for those whose function has been experimentally determined, blue when the function has been inferred by similarity and published, purple when the GT has been manually assigned based on visual inspection, and red identifies the suggestions proposed by the system based on the results provided by BLAST+ comparisons. (**B**) NMR chemical shifts and other parameters related to NMR experiments, non-sugar component (s) attached to the O-antigen structure, amino acid sequences of O-unit flippase and O-antigen polymerase, and literature data regarding the serogroup.

determined and published, (ii) blue when the function has been inferred by similarity searches and published, (iii) purple when the function has been manually assigned based on similarity searches and manual inspection but previously unpublished and (iv) the color red is used to indicate that the GT specificity (sugar donors and acceptors) was proposed by ECODAB based on amino acid and O-antigen structure comparison with GTs whose function is known (green, blue and purple).

In certain cases, these predictions are made for GTs that share similarity to a considerable number of other GTs, some of which may still have different functions. Moreover, the assignment of sugar donors and acceptors has to be consistent with the structure of the O-antigen. To overcome this situation, every GT in the system is accompanied by a list of candidate donor and acceptor sugars resulting from the most similar amino acid sequences found in the database. The abovementioned color code is also used to indicate the status of sugar donors and acceptors for each similar GT (Figure 3A). These colors may help to simplify the analysis of the results. For example, when pondering about different alternatives those with green color should be given a higher weighting.

Previous BLAST comparisons involved the manual inspection of a long list of results. The use of a stand-alone installation of BLAST+ to compare amino acid sequences has greatly improved this process since the results are generated in a fully automated way. The results from the sequence similarity search are ordered by decreasing *e*-value and the glycosidic linkages of the highest ranking GTs are compared with the target structure, when available, in order to locate mutual components. The prediction of GT function is then based on the highest ranking acceptor–donor pair with a structural similarity.

The following interface section, NMR, displays the parameters used to determine O-antigen structure by ^{1}H and ^{13}C NMR experiments, including the solvent of the compounds, the temperature of the sample, reference signal, pD values (where pD = pH + 0.4) and the resulting chemical shift data. The next part lists the non-sugar components linked to the O-antigen structure and the amino acid sequences of genes encoding O-unit flippase and O-antigen polymerase. The latter includes a link to the respective entry in the Uniprot Knowledgebase (Magrane and Consortium 2011). Finally, the literature section lists the most relevant bibliographic information about the particular serogroup in order to support the accuracy of the data; and if available, the PubMed entry is linked.

In addition to the main interface, ECODAB implements two ways to aid the exploration of entries: a summary page and a search option. In the summary page, a table is displayed. Every cell represents a serogroup, and inside the cells a label in the format aa/bb/cc is contained: aa indicates the number of GTs with known or manually assigned function, bb is the number of GTs for each entry and cc is the number of anticipated GTs based on the structure. This table is updated automatically with each new entry in the database. For every cell, a color code is used: (i) green for serogroups whose structures are stored in ECODAB, (ii) blue when the O-antigen structure does not exist in ECODAB, but there is GT information, (iii) red for those serogroups whose structures are not stored in ECODAB and (iv) black for serogroups that have been removed.

The search option facilitates the retrieval of entries that satisfy a set of rules based on substructure, chemical shifts, GT name and/or GENE-BANK code (Benson et al. 2011). All of the parameters manually typed in the interface are used to perform the search query. Accordingly, the more parameters that were used, the more defined the search would become. Substructure search might be the most usual query parameter. In this case, ECODAB will retrieve all entries in the

database where the given matches form part of the O-antigen structure. ^{1}H and ^{13}C NMR chemical shift data can be searched individually. The query values are restricted to two decimals and they must be separated by an empty space. Moreover, the user can select whether to retrieve entries that match exactly the query values or that lie within an interval of 0.05 ppm, for example, a query value of 2.0 will retrieve chemical shifts within 1.98 and 2.02. Finally, searches can also be performed by GT name and/or GENEBANK code.

Initially, only data on *E. coli* were collected in ECODAB. However, the new version of the application has been expanded to include data also on other bacteria. This change is particularly useful when the lack of resemblance to other GTs present in *E. coli* prevents the prediction of GT function. In these cases, other bacteria that produce analogous O-antigens and whose GT functions have been determined can support the prediction of GT functions for *E. coli*; for example, *Shigella* and *E. coli* are so closely related that some forms of *Shigella* are considered clones of *E. coli* (Lan and Reeves 2002). Hence, sequence comparison between different bacterial organisms can certainly be helpful. Data on other bacteria, for instance *S. pneumoniae*, are displayed on a similar layout as described for *E. coli*, but the interface is accessed separately in order to preserve the focus on *E. coli*.

Since its last iteration, ECODAB has grown to include data on 132 O-antigen structures and 338 GTs of an existing 180 O-antigens. A total of 188 GTs have their function predicted or experimentally determined, of which 94 were automatically predicted by ECODAB, 17 have been experimentally determined, 49 have been inferred and published and 28 were determined through similarity searches in this study (Figure 4). An example of the latter is WelM of *E. coli* O102 that shares a high-sequence similarity to WffH in O130 as well as an identical linkage, namely β-D-GalNAc-(1→3)-D-Gal, present in both structures. Neither of these GTs share significant sequence similarities with any other GT in ECODAB, and it is therefore concluded that they are responsible for the formation of the aforementioned linkage.

Many GTs share common origins and consequently similar GTs tend to cluster in groups. Being able to infer a specific GT function within a group, which shares sequence similarities where GT functions were previously unknown, gives ECODAB a reference point for making automatic predictions and often increased the number of predicted GT functions considerably. This is the case for the abovementioned pair in *E. coli* O130 and O102 as well as for WfcD of the O141 serogroup. In this study, 12 of the 94 GTs that were automatically
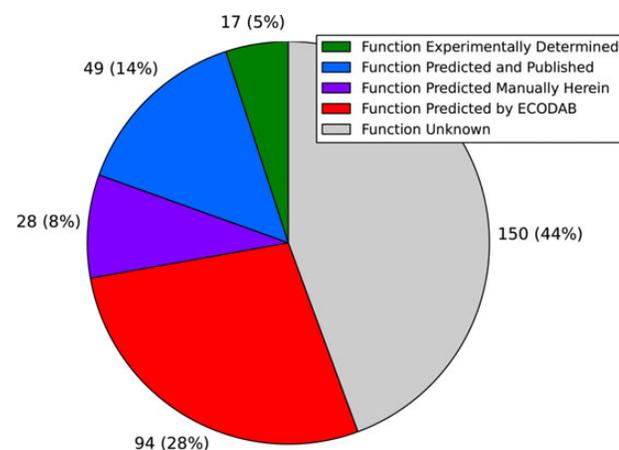

Fig. 4. The function of glycosyltransferases and relative occurrences as stored in ECODAB, based on 338 GT occurrences.

predicted by ECODAB belong to serogroups with an unknown structure and these should be able to help in future structure elucidations. The reduction in the number of GT functions determined by ECODAB compared with previous iterations is due to an increase of known published functions combined with stricter criteria for the automated predictions, rendering some previously inferred GT functions as unknown.

Currently, identification of putative GTs, based on genetic data, is becoming more common, but that is not the case for predictions of function of these proteins. The biochemical determination of GT function is challenging and only a small number of putative GTs has been studied biochemically. ECODAB takes advantage of the available data on GTs and uses this source to suggest the function of hitherto undetermined GTs by amino acid sequence similarity to GTs with known function. The results may not be conclusive, but they can be used to guide future experiments for determining O-antigen structures ([Fontana et al. 2014](#)).

The upgrade of ECODAB to implement a relational database has improved the overall performance of the application. Especially, beneficial is the concept of a Database Management System or DBMS ([Codd 1990](#)), which handles the way a relational database is created and how the stored data are organized, maintained, retrieved and manipulated. The DBMS also enforces the constraints that govern the relationships between tables and guarantee the integrity of the data by avoiding inconsistencies, duplication, absence, orphanage or obsolescence. Additionally, the DBMS along with the structure of a relational database comprise a flexible programming environment.

In conclusion, ECODAB can be freely accessed at http://www.casper.organ.su.se/ECODAB/ and together with the Carbohydrate Structure Database ([Toukach 2011](#); [Egorova and Toukach 2014](#)) facilitates the retrieval of structural information of O-antigen polysaccharides from *E. coli* in particular and from other bacteria in general.

Materials and methods

ECODAB has been further developed in order to replace the previous file-system database with a relational database managed by MySQL. The programs to fully automate storage, modification, recovery and display of information were written in PHP. ECODAB can be accessed through a standard Web browser, and additional software does not need to be installed. The database is continuously updated with published data on newly characterized serogroups, mainly from *E. coli*, but the possibility exists to expand the database to include other bacteria.

In order to make predictions on GT functions, amino acid sequence similarity to GTs with determined specificity can be used to suggest a function. In ECODAB, the amino acid sequence of each newly added GT is retrieved from the NCBI portal and stored in the database. The collection of sequences is subsequently used as an input file for a locally installed version of stand-alone BLAST+ ([Altschul et al. 1990](#)) in order to perform sequence comparisons.

The *e*-value is a statistical measure of the significance of the results of a BLAST search. A value of 0.001 for example, implies the probability of 0.1% that a given similarity arose by chance alone. The lower the value, the more significant the result. In ECODAB an *e*-value threshold of 0.001 is used when selecting similar GTs and then a bit score of 90 for the display of records.

Funding

Conflict of interest statement

None declared.

Abbreviations

ECODAB, *Escherichia coli* O-antigen database; GTs, glycosyltransferases; LPS, lipopolysaccharide; O-PS, O-antigen polysaccharide.

References

Allison GE, Verma NK. 2000. Serotype-converting bacteriophages and O-antigen modification in *Shigella flexneri*. *Trends Microbiol*. 8:17–23.

Altschul SF, Gish W, Miller W, Myers EW, Lipmap DJ. 1990. Basic local assignment tool. *J Mol Biol*. 215:403–410.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2011. GenBank. *Nucleic Acids Res*. 39:D32–D37.

Codd E. 1990. *The Relational Model for Database Management*. 2nd ed. Reading: Addison-Wesley Publishing Company.

Coutinho PM, Rancurel C, Stam M, Bernard T, Couto FM, Danchin EGJ, Henrissat B. 2009. Carbohydrate-active enzymes database: principles and classification of glycosyltransferases. In: von der Lieth C-W, Lütteke T, Frank M, editors. *Bioinformatics for Glycobiology and Glyconomics: An Introduction*. West Sussex: John Wiley & Sons. p. 91–118.

DebRoy C, Roberts E, Fratamico PM. 2011. Detection of O antigens in *Escherichia coli*. *Anim Health Res Rev*. 12:169–185.

Egorova KS, Toukach PV. 2014. Expansion of coverage of Carbohydrate Structure Database (CSDB). *Carbohydr Res*. 389:112–114.

Ewing WH. 1986. The genus *Escherichia*. In: Edwards PR, Ewing WH, editors. *Edwards & Ewing's Identification of Enterobacteriae*. 4th ed. New York: Elsevier Science. p. 93–134.

Fontana C, Lundborg M, Weintraub A, Widmalm G. 2014. Rapid structural elucidation of polysaccharides employing predicted functions of glycosyltransferases and NMR data: Application to the O-antigen of *Escherichia coli* O59. *Glycobiology*. 24:450–457.

Guo H, Kong Q, Cheng J, Wang L, Feng L. 2005. Characterization of the *Escherichia coli* O59 and O155 O-antigen gene clusters: The atypical wzx genes are evolutionary related. *FEMS Microbiol Lett*. 248:153–161.

Hug I, Couturier MR, Rooker MM, Taylor DE, Stein M, Feldman MF. 2010. *Helicobacter pylori* lipopolysaccharide is synthesized via a novel pathway with an evolutionary connection to protein N-glycosylation. *PLoS Pathog*. 6:e1000819.

Kalynych S, Ruan X, Valvano MA, Cygler M. 2011. Structure-guided investigation of lipopolysaccharide O-antigen chain length regulators reveals regions critical for modal length control. *J Bacteriol*. 193:3710–3721.

Kaper JB, Nataro JP, Mobley HL. 2004. Pathogenic *Escherichia coli*. *Nat Rev Microbiol*. 2:123–140.

Lan R, Reeves PR. 2002. *Escherichia coli* in disguise: Molecular origins of *Shigella*. *Microbes Infect*. 4:1125–1132.

Li X, Perepelov AV, Wang Q, Senchenkova SN, Liu B, Shevelev SD, Guo X, Shashkov AS, Chen W, Wang L, et al. 2010. Structural and genetic characterization of the O-antigen of *Escherichia coli* O161 containing a derivative of a higher acidic diamino sugar, legionaminic acid. *Carbohydr Res*. 345:1581–1587.

Linnerborg M, Weintraub A, Widmalm G. 1999a. Structural studies utilizing ^{13}C-enrichment of the O-antigen polysaccharide from the enterotoxigenic *Escherichia coli* O159 cross-reacting with *Shigella dysenteriae* type 4. *Eur J Biochem*. 266:246–251.

Linnerborg M, Weintraub A, Widmalm G. 1999b. Structural studies of the O-antigen polysaccharide from the enteroinvasive *Escherichia coli* O164 cross-reacting with *Shigella dysenteriae* 3. *Eur J Biochem*. 266:460–466.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*. 42:D490–D495.

Lundborg M, Modhukur V, Widmalm G. 2010. Glycosyltransferase functions of *E. coli* O-antigens. *Glycobiology*. 20:366–368.

Lycknert K, Widmalm G. 2004. Dynamics of the *Escherichia coli* O91 O-antigen polysaccharide in solution as studied by carbon-13 NMR relaxation. *Biomacromolecules*. 5:1015–1020.

Magrane M, Consortium U. 2011. UniProt Knowledgebase: A hub of integrated protein data. *Database (Oxford)*. 2011:bar009.

Nataro JP, Kaper JB. 1998. Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev*. 11:142–201.

Ørskov F, Ørskov I. 1992. *Escherichia coli* serotyping and disease in man and animals. *Can J Microbiol*. 38:699–704.

Ørskov I, Ørskov F, Jann B, Jann K. 1977. Serology, chemistry, and genetics of O and K antigens of *Escherichia* coli. *Bacteriol Rev*. 41:667–710.

Prager R, Strutz U, Fruth A, Tschäpe H. 2003. Subtyping of pathogenic *Escherichia coli* strains using flagellar (H)-antigens: Serotyping versus *fliC* polymorphisms. *Int J Med Microbiol*. 292:477–486.

Rocchetta HL, Burrows LL, Pacan JC, Lam JS. 1998. Three rhamnosyltransferases responsible for assembly of the A-band D-rhamnan polysaccharide in *Pseudomonas aeruginosa*: a fourth transferase, WbpL, is required for the initiation of both A-band and B-band lipopolysaccharide synthesis. *Mol Microbiol*. 28:1103–1119.

Samuel G, Reeves P. 2003. Biosynthesis of O-antigens: Genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr Res*. 338:2503–2519.

Stenutz R, Weintraub A, Widmalm G. 2006. The structures of *Escherichia coli* O-polysaccharide antigens. *FEMS Microbiol Rev*. 30:382–403.

Stevenson G, Dieckelmann M, Reeves PR. 2008. Determination of glycosyltransferase specificities for the *Escherichia coli* O111 O antigen by a generic approach. *Appl Environ Microbiol*. 74:1294–1298.

Toukach PV. 2011. Bacterial carbohydrate structure database 3: Principles and realization. *J Chem Inf Model*. 51:159–170.

Urbina F, Nordmark E-L, Yang Z, Weintraub A, Scheutz F, Widmalm G. 2005. Structural elucidation of the O-antigenic polysaccharide from the enteroaggregative *Escherichia* coli strain 180/C3 and its immunochemical relationship with *E. coli* O5 and O65. *Carbohydr Res*. 340: 645–650.

Valvano MA, Furlong SE, Patel KB. 2011. Genetics, biosynthesis and assembly of O-antigen. In: Knirel YA, Valvano MA, editors. *Bacterial Lipopolysaccharides*. Wien: Springer-Verlag. p. 275–310.

Wang L, Liu B, Kong Q, Steinrück H, Krause G, Beutin L, Feng L. 2005. Molecular markers for detection of pathogenic *Escherichia coli* strains belonging to serogroups O138 and O139. *Vet Microbiol*. 111:181–190.

Wang W, Perepelov AV, Feng L, Shevelev SD, Wang Q, Senchenkova SN, Han W, Li Y, Shashkov AS, Knirel YA, et al. 2007. A group of *Escherichia coli* and *Salmonella enterica* O antigens sharing a common backbone structure. *Microbiology*. 153:2159–2167.