EK3D : An *E. coli* K antigen 3D Structural Database

Kunduru Bharathi Reddy

A Dissertation Submitted to

Indian Institute of Technology Hyderabad

In Partial Fulfillment of the Requirements for

The Degree of Master of Technology

Department of Biotechnology

June, 2015

Declaration

I declare that this written submission represents my ideas in my own words, and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from the sources that have thus not been properly cited, or from whom proper permission has not been taken when needed.

K. Bharathi Reddy

(Signature)

(Kunduru Bharathi Reddy)

(Bo13m1010)

Approval Sheet

This thesis entitled "An *E. coli* K antigen 3Dimensional Database" by Kunduru Bharathi Reddy is approved for the degree of Master of Technology from IIT Hyderabad.

Dr. Anindya Roy
Assistant Professor
Department of Biotechnology, IIT Hyderabad
Internal examiner

Dr. Amirtham Rajagopal
Assistant Professor
Department of Civil Engineering, IIT Hyderabad
External examiner

Dr. Thenmalarchelvi Rathinavelan
Assistant Professor
Department of Biotechnology, IIT Hyderabad
Adviser

Acknowledgements

I would like to express my deep gratitude to my guide **Dr Thenmalarchelvi Rathinavelan.**

I thank my lab members **Sanjana, Abhishek, Narender, Venkat and Yogeesh** for their co-operation & support.

I wish to thank **Goutham, Ravinder** and others who helped me to successfully complete my project work.

I would also like to thank IITH for its support.

Thank you very much.

Dedicated to

I would like to dedicate my work to my beloved parents, family & my guide Dr. Thenmalarchelvi Rathinavelan, who helped me to successfully complete the project.

Abbreviations

MDR-Multi Drug Resistance

WHO-World Health Organization

AMR-Anti Microbial Resistance

MRSA-Methicillin Resistant Staphylococcus

CPS-Capsule polysaccharide

LPS-Lipopolysaccharide

EPS-Exopolysaccharide

K_{LPS}-Capsular antigen linked to LPS lipid A core

ABC transporter- ATP-binding cassette transporters

GT-Glycosyl transferase

GalF-Galactose locus

UndPP- Undecaprenyl pyro phosphate

Kps-Group-2 & 3 locus

Ser A-Serine A locus

Kdo-Diacyl glycerol phosphate

CSDB-Carbohydrate structure database

PolysacD -Database of antigenic polysaccharides of microbes

ECODAB-*Escherichia coli* O-antigen database

GFDB-Glycan fragment database

EcoCyc- *Escherichia coli* K-12 MG1655

BPGD-Bacterial polysaccharide gene database

HTML: Hyper Text Markup Language

LAMP-(Linux+Apache2+mysql+Php)

PHP: Hyper text pre processor

SQL-Structure based query language

DBMS-Database management system

URL-Uniform resource locator

PDB-Protein data bank

CATH- Protein Structure Classification Database

Abstract

Diseases caused by multidrug resistant Gram-negative bacterial strains such as Enterobacteriaceae members (*Salmonella, Pseudomonas & Klebsiella*) claim millions of lives every year. One of the major virulence determinants of these bacteria are the polymeric surface antigens like capsular polysaccharides (CPS), exopolysaccharides (EPS) & lipopolysaccharides (LPS). A detailed understanding of these surface antigens is essential to develop drugs against bacterial infections. However, this is made difficult by the large variety of surface antigens that have been discovered to date. To this end, an organized repository of all available structures, functions and biochemical information about the various Gram-negative surface antigens is highly essential. Although several efforts have been made previously to develop databases of bacterial surface antigens, they offer only limited information. In the present study, we focus on the development of a database for Group capsular polysaccharides (K antigens) of *E. coli*. It is noteworthy that *E. coli* is used as a model system here, as it is very well studied and its capsular polysaccharides have been well characterized. The database, named EK3D (*E. coli* K antigen 3D Structural Database) provides information about the structure of Group capsular antigens and the proteins involved in its biosynthesis. The database will also enable generation of polymeric CPS structures of varying lengths for modeling and Molecular dynamics (MD) studies. Thus, EK3D will provide comprehensive information about group K antigens of *E. coli* and be a valuable resource for researchers.

Contents

4 List of Figures

5 List of Tables

CHAPTER 1

INTRODUCTION

1.1 BACTERIA AND ASSOCIATED DISEASES

Microorganisms play an important role in causing a wide range of diseases. Bacteria are one among these that exhibit higher rate of antimicrobial resistance, and are known to contain structure like cell envelope that helps in protection and adaptation of the organism. The cell envelope of Gram positive organisms has thick layer of peptidoglycan. Teichoic acids, a type of polyalcohols are present and are linked to lipids to form lipoteichoic acids within the cytoplasmic membrane. In Gram negative cell envelope peptidoglycan layer is thin, teichoic acids are absent and there is an outer membrane in addition to peptidoglycan layer which is made up of phospholipids and lipopolysaccharides. The porins present in the impermeable phospholipid membrane allow passive transport of many ions, sugars and aminoacids across the outer membrane. Periplasm, present between inner and outer membrane, contains peptidoglycan layer and is involved in transport of proteins. Both Gram positive and Gram negative organisms are known to be responsible for causing wide range of diseases.

Gram positive organisms and their associated diseases include *Staphylococcus* and *Streptococcal* infections. Gram negative organisms like *E. coli, Klebsiella, Neisseria and Haemophilus* are known to be associated with diarrhoea, neonatal meningitides, urinary tract infections, septicemia etc. Gram negative and Gram positive bacteria and their associated diseases, no of capsular serotypes and their antigenic determinant are shown in Table 1. Though the diseases caused by these bacteria can be treated with antibiotics, the overuse of these antibiotics is one of the reasons for the development of resistance.

1.2 Multi Drug Resistance

Emergence of multi drug resistance (MDR) has become a ‘major global threat’, claiming millions of lives every year. Currently the estimate is 10 million deaths each year (Reardon, 2014). The steadily rising MDR strains of bacteria (Reardon, 2014), as indicated in Figure 1,

has essentially brought us to a "post antibiotic era", as described by WHO (World Health Organization). Hence, bacterial strains that have evolved multi drug resistance, such as *Mycobacterium, Klebsiella, Gonorrhoea* etc are major threat to public health. The present data where in, the action of major drug like Carbapenem is the last resort drug to treat disease

Table 1. Clinically important bacteria & associated diseases (Nwodo et al., 2012)

Bacterial Species	Pathogenic serotype	Capsular Antigen	Associated clinical Disease
E. coli	>80	K antigen	Diarrhoea, neonatal meningitis and urinary tract infection
N. meningitidis	>10	K antigen	Meningitis, meningococcemia
K. pneumonia	>80	K antigen	Pneumonia, bacteremia, urinary tract infection (UTI)
St. pneumoniae	>96	CPS	Otitis media, bronchopneumonia and meningitis
S. aureus	>11	CPS	Staphylococcal scalded skin syndrome, septic arthritis, staphylococcal endocarditis and atopic dermatitis

pneumonia no longer works. Bacteria mutate rapidly and are capable of easily evading the power of antibiotics. The term ESKAPE was given to collectively summarize organisms *Enterococcus, Staphylococcus, Klebsiella, Acinetobacter, Pseudomonas*, and ESBL (*Enterobacter* and *Esherichia coli*) that deveoped multidrug resistance.

O'Neill putforth the term BRIC (Brazil, Russia, India and China) and more recently MINT (Mexico,Indonesia, Nigeria and Turkey) states where the impact of multi drug resistance varies (Reardon, 2014). Gram negative bacteria are at the growing level of drug resistant infections. Deaths attributed to AMR (antimicrobial resistance) every year by 2050 will be more than one in four deaths affecting great number of population. Current research suggests that switching

between two antibiotics in a well designed sequence may prove it to be a new way to fight the antimicrobial resistance for which clinical trials are on the way (Westly, 2012).

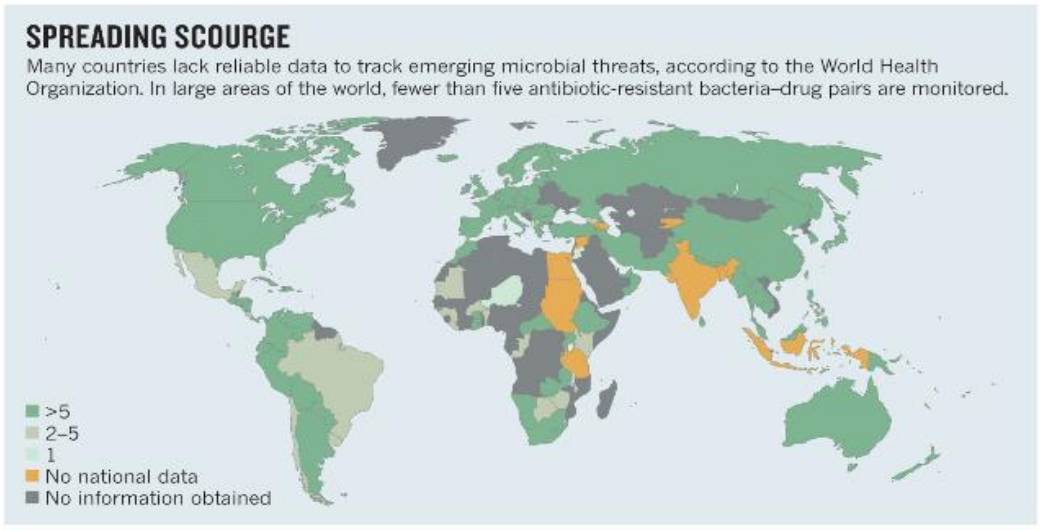


Figure 1. Global map illustrating the world wide spread of antibiotic resistance (Reardon, 2014). The antibiotic-resistant bacteria and drug pairs monitored were fewer than 5 in large number of areas and are shown in dark green colour. No data is available in some areas as shown in orange.

Recent WHO report on global surveillance shows (Figure 2 & 3) increased resistance rate towards broad spectrum antibiotics. Antibacterial resistance in *E. coli* against third generation cephalosporins & carbapenems is due to the release of beta-lactamases which destroy beta-lactam antibacterial drugs.

WHO South East Asia region report

Information and data available for selected diseases and organisms reveal that AMR is a burgeoning and often neglected problem. Since 2011 the awareness has been increasing to combat AMR.

Resistance to cephalosporins and fluoroquinolones in *Escherichia coli*

Data obtained from various states has been summarized by WHO and the (Figure 3) shows the information collected from different regions summarizing the resistance rate to be higher for fluoroquinolones. Some bacterial strains emerge as antibiotic resistant strains upon acquiring drug resistance conferring pathogenicity to the organism. Numerous studies on bacteria have focused on the cell surface glycoconjugates of microorganisms, because of their importance in pathogenicity.

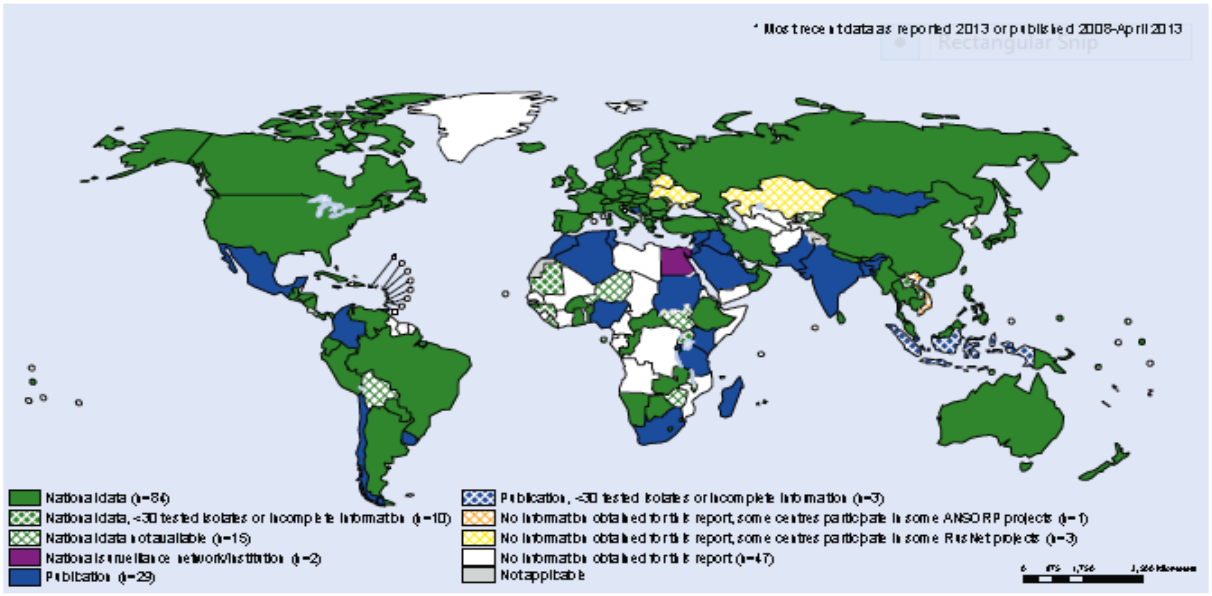


Figure 2. Global map on cephalosporin resistance (Review on Antimicrobial resistance, Global surveillance, 2014).

Table 2. Data obtained from Member States as summarized by WHO on *E. coli* MDR (Review on Antimicrobial resistance, Global surveillance, 2014).

Drug	Data sources based on at least 30 tested isolates	Overall reported range of resistant proportion (%)
Fluoroquinolones	South-East Asia Region – National data (n=5 countries) – Publications (n=19) from 2 additional countries	32–64 4–89

Cephalosporins	South-East Asia Region – National data (n=5 countries) –Publications (n=26) from 2 additional countries	16–68 19–95

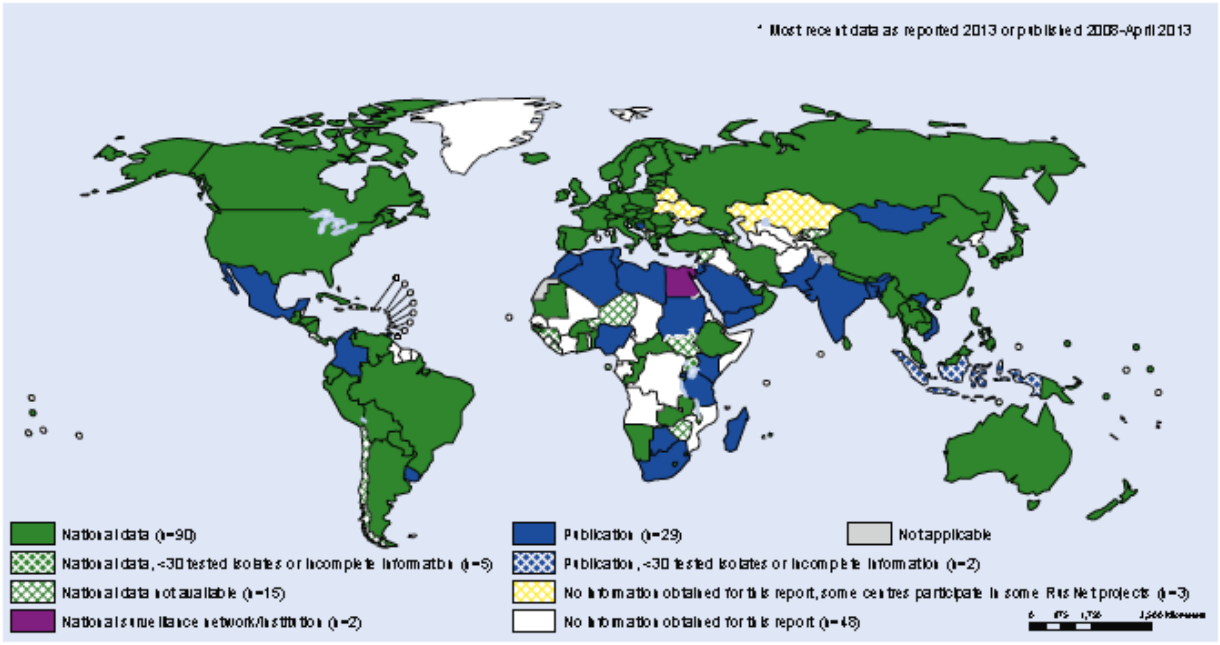


Figure 3. Global map on fluoroquinolone resistance (Review on Antimicrobial resistance, Global surveillance, 2014).

1.3 BACTERIAL POLYSACCHARIDES

Gram-negative bacteria are known to contain polysaccharides as the abundant polymer and there is a remarkable variation in the polysaccharides produced by various strains.

1.3.1 Diversity of bacterial polysaccharides

Polysaccharides are one of the virulent factors in Gram negative bacteria. The major polysaccharides known are lipopolysaccharides and capsular polysaccharides. Different O antigens and K antigens are reported in various serotypes of *E. coli*. Nearly 167 O serogroups and 80 K serogroups are noted. Another exopolysaccharide that is loosely associated with cell

present in many *E. coli* strains is colonic acid. Although this doesn't play a role in bacterial pathogenecity, it helps bacterium to survive outside the host.

The antigen serotyping is useful in understanding the pathogenesis and in epidemiological studies in tracing the origin of disease. Each antigenic structure composition is different for a particular serotype which helps in identification of specific strains. Based on the antigens associated with the outer membrane like LPS (O antigen), CPS (K antigen) and flagellar (H antigen), in 1940's Kauffman proposed earliest classification based on O, K and H antigens. Wherein, O antigen consists of polysaccharide attached to a lipid A-core unit of outer membrane (part of LPS). H antigen is flagellar subunit associated antigen which is also a determinant of virulence. K antigen is the high molecular weight capsular polysaccharide associated with outer membrane of the bacteria (Figure 4). There exists a greater structural diversity in polysaccharides between different bacterial species and also within the same bacterial species which can be exploited to engineer novel polysaccharides molecules with particular biochemical or immunological properties. Significant advances has been made in understanding the biosynthesis and surface assembly of these structures (Refer complex carbohydrate structure database-www.ccrc.uga.edu/) (Whitfield, 1988).

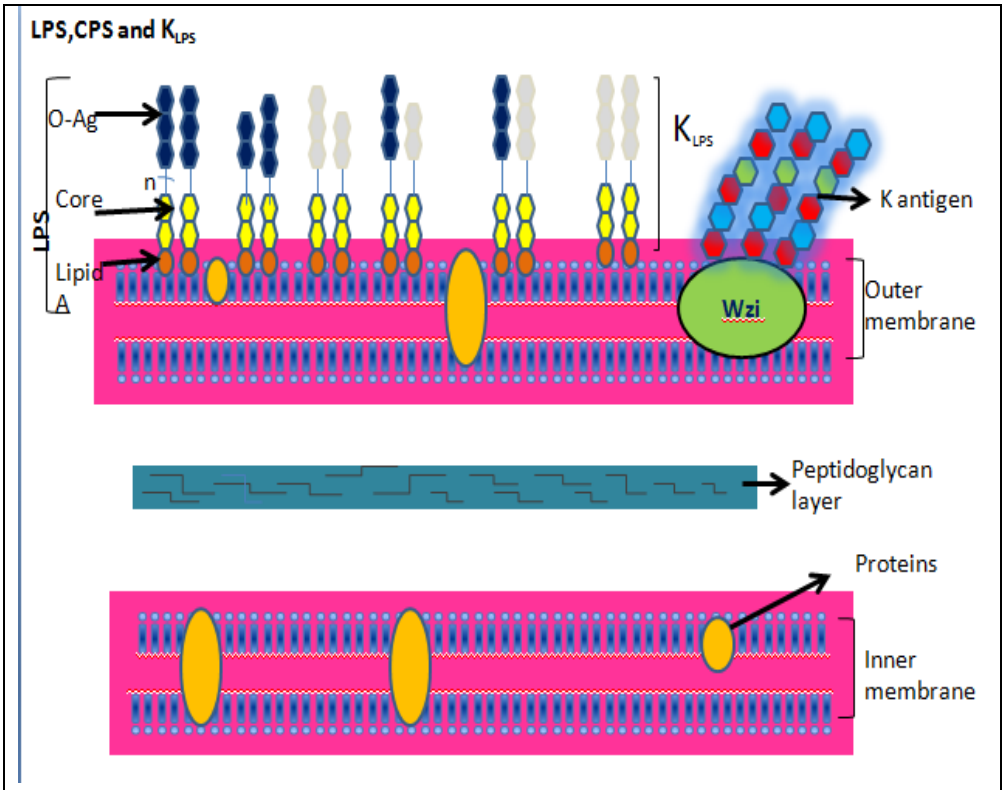


Figure 4. Schematic representation of bacterial LPS and CPS showing surface associated Gram negative polysaccharides. The K antigen is shown attached to the lectin wzi and O antigen is shown attached to core and lipid A of LPS and K_{LPS} is K antigen attached to LPS**.**

1.3.2 Capsular antigen/ K antigen

Capsule is mainly, a polysaccharide layer present in both Gram positive and Gram negative organisms. Figure 5 depicts the electron micrograph of *E. coli* capsule showing the morphology of CPS on surface. Capsular polysaccharide (CPS) is tightly associated with cell surface non-covalently and protects bacteria from host immune system. CPS serves as an important model system for studies on other bacteria.

Functions of capsules include prevention of dessication, adherence, resistance to specific (mimic host cellular proteins) & non-specific host immunity (complement-mediated opsonophagocytosis).

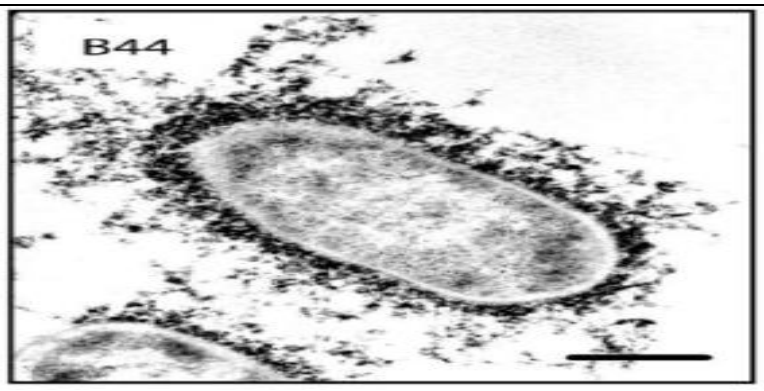


Figure 5. Electron micrograph illustrating the CPS on bacterial surface of *E. coli B44 (O9:K30)*. Adopted from (Rahn et al., 2003).

1.4 *E. coli* CAPSULAR GROUP CLASSIFICATION

Considering the present *E. coli* model system, classification of *E. coli* into groups is based on K antigen, its biosynthesis mechanism, genetics & biochemical characters. Based on capsular antigen surface expression & assembly mechanism, there are only two pathways used in *E. coli*

i.e, Wzy-dependent pathway and ABC transporter pathway. *E. coli* group 1 and 4 capsules share a common assembly system, and this is different from the one used for group 2 and 3 capsules.

Table 3. *E. coli* capsular group classification showing different characteristics (Whitfield, 2006)

Characteristics	Group-1	Group-2	Group-3	Group-4
Type of polysaccharide	Acidic polysaccharides (Uronic acids)	Repeat units vary extensively	Repeat units vary	Acidamido sugars
Occurrence	Found in isolates that cause intestinal infections	Found in isolates that cause extraintestinal infections	Found in isolates that cause extraintestinal infections	Found in isolates that cause intestinal Infections
Co-expression with colonic acid	Absent	Present	Present	Present
Terminal lipid moiety	KLPS-Lipid A KAg-unknown	α-glycerophosphate	α-glycerophosphate?	KLPS-Lipid A KAg-unknown
Polymerization type	Wzy dependent	ABC transporter dependent	ABC transporter dependent	Wzy dependent
Serotypes	K27,K29,K30	K1,K2a,K4,K5	K10,K54	K40
Capsular similarity	*Klebsiella, Erwinia*	*Neisseria, Haemophilus*	*Neisseria, Haemophilus*	None

Group 1 & 4 and group 2 & 3 have similarities with respect to genetic locus, thermostability, mechanism of biosynthesis (Table 3). Group-1 capsules are high molecular weight polysaccharides with low electrophoretic mobility and contain hexouronic acids as the acidic component. They are coexpressed with essentially O8 and O9 antigens at all growth temperatures, and are heat stable. Group-II capsules on the other hand are known to contain acidic components like KDO, NeuNAc or phosphate as unusal residues. They show greater

electrophoretic mobility and resemble *Neisseria meningitides or Haemophilus influenzae.* These capsules are heat labile at P^H 5-6 and are coexpressed with many O antigens.

Table 4. List of K antigens belonging to different *E. coli* group capsules.

Group-1- K26, K27, K28, K29, K30 ,K31, K32, K33, K34, K35, K36, K37, K39, K42, K55, K102, K103

Group-2- K1, K2a, K2ab, K3, K4, K5, K6 , K7, K11, K12, K13, K14, K15, K16, K18, K19, K20, K22, K23, K24, K51, K52, K53, K54, K74, K92, K93, K95, K97, K100

Group-3- K3, K10, K11, K54, K96, K98

Group-4- K8, K9, K38, K40, K44, K45, K46, K47, K48, K49, K50, K57, K83, K85, K87, O111, K101

(Max. Sussman, 1997, Escherischia coli : Mechanisms of Virulence, Pg-113-122, books.google.co.in/ ,Orskov et al, Methods in microbiology,Vol-14).

Depending on the sugar composition, linkage specificity, as well as substitution with non carbohydrate residues about 80 K antigen serotypes are known in *E. coli.* The structures of many of these antigens is already determined. However, no 3D structures are available.

1.4.1 **Biosynthesis & Surface assembly of Group-1 to Group-4 capsules**

1.4.1.1 Group-1 capsules biosynthesis and surface assembly

Bacteria use limited biosynthesis mechanisms despite diversity in capsular phenotypes. The biosynthetic pathway and the proteins involved in biosynthesis reflect the unique repeat unit structure of each K antigen. Group 1 K antigens are expressed as CPS and K_{LPS}. In case of CPS, the repeat unit polymerizes to form a heavy molecular weight structure that non-covalently associates with cell surface whereas K_{LPS} consists of K antigen repeating unit linked to lipid A core of LPS. Gram-negative bacterial outer membrane has been represented in Figure 4.

Capsular assembly via Wzy-dependent pathway begins in the cytoplasm with the synthesis of lipid-linked polysaccharide repeat units. Initially first sugar is added by the initiating glycosyl transferase (WbaP) to the polyisoprenoid lipid carrier molecule undecaprenyl phosphate (UndP)

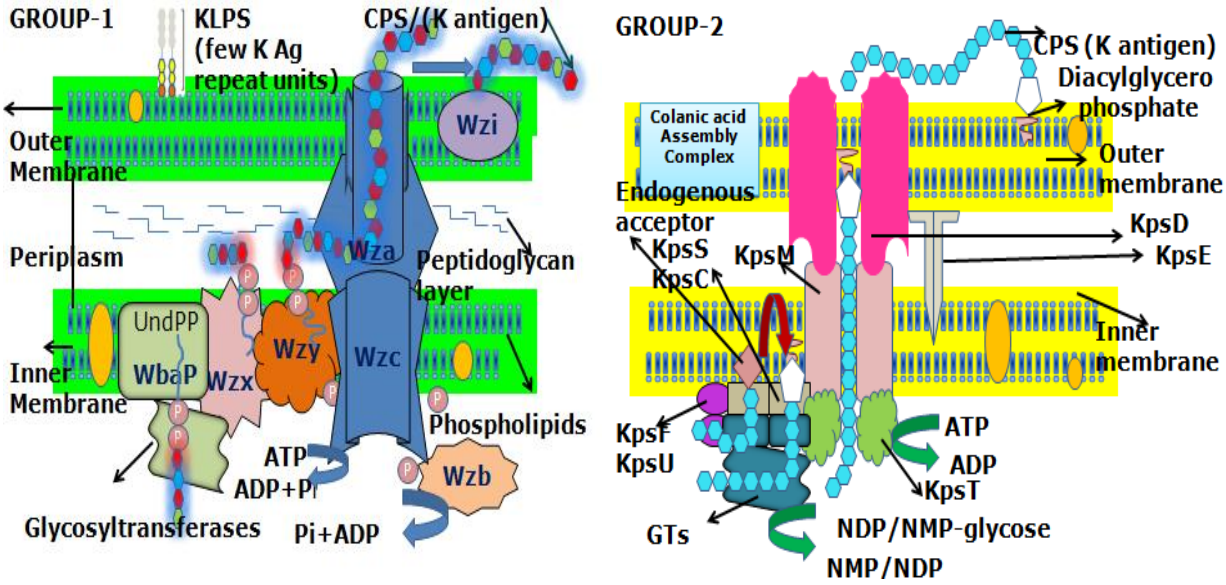


Figure 6. Schematic representation of Gram-negative bacterium capsular biosynthesis & its surface expression. Left view- Wzy dependent Polymerization pathway of group-1 and right view ABC transporter pathway of group-2 capsules. LPS consists of O-antigen (Black) repeating unit covalently linked to lipid A (Orange), Core(Yellow). K_{LPS} consists of few K antigen repeat units (white) linked to lipid A core. K antigen (repeat unit with variable colours red, blue, green) non covalently attached to lectin protein Wzi.

in the form of an activated nucleotide precursor at the inner leaflet of the inner membrane (IM), (Undecaprenol also serves as a scaffold for organizing proteins including GTs (Glycosyl transferases), and the lipid: protein complexes may alter the biophysical properties of the local membrane environment, perhaps aiding the flipping activity) resulting in a pyrophosphate linkage with the carrier (undecaprenyl pyrophosphate, UndPP). Subsequent sugar addition occurs by specific glycosyltransferases which result in the synthesis of an individual UndPP linked polysaccharide repeat unit (referred to as an 'O unit), after which a series of inner membrane assembly proteins take over (Table 5). The UndPP-linked repeat units are translocated from the inner to the outer membrane by the Wzx flippase. Polymerization by Wzy via catch-and-release mechanism occurs. In case of K-Antigen synthesis sugar-unit addition occurs at the reducing terminus of the growing chain, the length of which is regulated by the polysaccharide co-polymerase. Continuation of polymerization requires Wzc phosphorylation (C-terminal tyrosine

residues) and successive dephosphorylation of Wzb (Protein tyrosine phosphatase) continues the polymerization cycle. Wzc is also involved in closing and opening of Wza. The polymerized glycan is then anchored onto the bacterial surface by Wzi that is located on the outer membrane. In K_{LPS} the K-antigen is found attached to the mature LPS molecule (Whitfield, 2006).

Table 5. Proteins involved in biosynthesis of group-1 capsular polysaccharides in *E. coli.*

PROTEIN	FUNCTION
WbaP, WbaZ, WbaO, WbaN	Glycosyl transferase
Wzx	Repeat unit exporter
Wzy	Repeat unit polymerase
Wza	Putative outer membrane lipo protein
Wzb	Putative acid phosphatase
Wzc	Putative trans membrane protein kinase
Wzi	Outermembrane lectin protein

1.5 DATABASES

Great number of databases have been providing information about carbohydrates which are often, referred to as glycans (Campbell et al., 2014). Although in the past, number of databases have been developed with respect to carbohydrate structures like, CSDB (Carbohydrate structure database), PolysacDB (Database of antigenic polysaccharides of microbes) (Aithal et al., 2012), ECODAB (*E. coli* O antigen database) (Rojas-Macias et al., 2015) etc, the information provided lacks the 3D structural data. ECODAB database provides information on O antigen structures, NMR chemical shifts, few cross reactivity relationships of a particular serogroups. Some databases give sequence based information like Glycan fragment database (GFDB) (Jo and Im,

2013), primary structure information, or glycoprotein structural information (GlycosuiteDB) (Cooper et al., 2003, Cooper et al., 2001). *E. coli* K antigens exhibit conformational complexity. EK3D was developed in order to provide the necessary information about the K antigen structures among different *E. coli* groups which enables the better understanding of host-pathogen interaction (Sarkar and Perez, 2012). EK3D may serve as an example to develop other antigen determinant based databases that are limited. The main features of EK3D are it offers different K antigen structures of *E. coli* group capsules, serves as a tool to generate polymeric structures of K-antigen repeat units, provides related database links for cross reference and its implementation as webserver through HTML coding provides data access for the research purpose.

SCOPE OF THE PRESENT STUDY

With the increased incidence of multi-drug resistant Gram-negative bacterial infections and the lack of effective treatment options, the focus of current research has shifted to the discovery of new antibacterial drug targets. This in turn requires an in-depth understanding of the bacterial surface antigen structure and underlying virulence mechanisms. In this study, we focus on one of the major Gram-negative virulence determinants, the Capsular polysaccharide. Though the CPS has been well characterized in a number of Gram-negative bacteria, a central repository of all available information is lacking. For instance, the ECODAB database (http://www.casper.organ.su.se/ECODAB) focuses on O-antigen structures of *E. coli* , EcoCyc (http://ecocyc.org) is restricted to only *E. coli* K12 strain. Likewise, BPGD gives information about bacterial polysaccharide genes but is restricted to nomenclature only (Rojas-Macias et al., 2015). Thus, a comprehensive K antigen database if present would be a valuable resource for researchers. In the present study, we develop EK3D (*E. coli* K antigen 3dimensional Structural Database), which provides information about Group-1 to 4 K antigen structures of *E. coli*. Structural information about the proteins involved in the assembly of the Group capsules is also available. We also aim to develop a modeling tool to generate Capsular polysaccharides of specified length from the known repeating unit structures. This database will thus be a valuable research tool that will help accelerate research in the field of capsular polysaccharides like host-pathogen interaction, docking studies and K antigen structure determination.

REFERENCES

AITHAL, A., SHARMA, A., JOSHI, S., RAGHAVA, G. P. & VARSHNEY, G. C. 2012. PolysacDB: a database of microbial polysaccharide antigens and their antibodies. *PLoS One,* 7**,** e34613.

CAMPBELL, M. P., RANZINGER, R., LUTTEKE, T., MARIETHOZ, J., HAYES, C. A., ZHANG, J., AKUNE, Y., AOKI-KINOSHITA, K. F., DAMERELL, D., CARTA, G., YORK, W. S., HASLAM, S. M., NARIMATSU, H., RUDD, P. M., KARLSSON, N. G., PACKER, N. H. & LISACEK, F. 2014. Toolboxes for a standardised and systematic study of glycans. *BMC Bioinformatics,* 15 Suppl 1**,** S9.

COOPER, C. A., HARRISON, M. J., WILKINS, M. R. & PACKER, N. H. 2001. GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources. *Nucleic Acids Res,* 29**,** 332-5.

COOPER, C. A., JOSHI, H. J., HARRISON, M. J., WILKINS, M. R. & PACKER, N. H. 2003. GlycoSuiteDB: a curated relational database of glycoprotein glycan structures and their biological sources. 2003 update. *Nucleic Acids Res,* 31**,** 511-3.

JO, S. & IM, W. 2013. Glycan fragment database: a database of PDB-based glycan 3D structures. *Nucleic Acids Res,* 41**,** D470-4.

Max. Sussman, 1997, Escherischia coli : Mechanisms of Virulence, Pg-113-122, books.google.co.in/ ,Orskov et al, Methods in microbiology,Vol-14).

NWODO, U. U., GREEN, E. & OKOH, A. I. 2012. Bacterial exopolysaccharides: functionality and prospects. Int J Mol Sci, 13, 14002-15.

RAHN, A., BEIS, K., NAISMITH, J. H. & WHITFIELD, C. 2003. A novel outer membrane protein, Wzi, is involved in surface assembly of the Escherichia coli K30 group 1 capsule. *J Bacteriol,* 185**,** 5882-90.

REARDON, S. 2014. Antibiotic resistance sweeping developing world. Nature, 509, 141-2.

ROJAS-MACIAS, M. A., STAHLE, J., LUTTEKE, T. & WIDMALM, G. 2015. Development of the ECODAB into a relational database for Escherichia coli O-antigens and other bacterial polysaccharidesdagger. *Glycobiology,* 25**,** 341-7.

SARKAR, A. & PEREZ, S. 2012. PolySac3DB: an annotated data base of 3 dimensional structures of polysaccharides. *BMC Bioinformatics,* 13**,** 302.

WESTLY, E. 2012. India moves to tackle antibiotic resistance. *Nature,* 489**,** 192.

WHITFIELD, C. 1988. Bacterial extracellular polysaccharides. *Can J Microbiol,* 34**,** 415-20.

WHITFIELD, C. 2006. Biosynthesis and assembly of capsular polysaccharides in Escherichia coli. *Annu Rev Biochem,* 75**,** 39-68.

WESTLY, E. 2012. India moves to tackle antibiotic resistance. Nature, 489, 192.

http://www.casper.organ.su.se/ECODAB)

http://ecocyc.org

https://books.google.co.in/books?hl=en&lr=&id=7vFpeDcjBH0C&oi=fnd&pg=PA145&dq=Hull+1997+Jann+%26+Jann++E.coli+capsular+polysaccharides&ots=fNUe0yx77O&sig=R7-yXQjoHmgglvxjr_rPW_LG8uM#v=onepage&q&f=false page-145-161

CHAPTER 2

METHODS

Bacteria produce larger amounts of glycans as part of capsular polysaccharide, lipopolysaccharide, EPS and antibiotic glycosides. There is a greater need to understand the structure, recognition, metabolism and biosynthesis of glycans and glycoconjugates. In most of the pathogens, CPS are the well known virulent factors (Merritt et al., 2013). As *E. coli* cells serve as model organism to study glycosylation and glycoprotein expression developing a database with a systamatic record of related information may serve as a major tool in making therapeutics and vaccines. An organized database is easy to track and verify and thus, database of *E. coli* K-antigens can store related information on different subjects. In this context, we aim to create a 3D database of *E. coli* K antigen structures, named EK3D (*E. coli* K antigen 3D Structural Database). A basic overview of database development is provided below, followed by a detailed description of EK3D.

2.1 DATABASE DEVELOPMENT

Database management systems are the computer based software applications that interact with the user to analyze and capture data. The data is initially entered and stored for the retrieval process. On basic linux ubuntu server the following drives were mounted (MYSQL, APACHE2, PHP) (server sided).

2.1.1 Mounting drives

The following drives were installed to establish a database.

Step by step process to install programmes like Apache2, Mysql, PHP, Kozomer was carried out. Popular setups for ubuntu server are designated as LAMP installations (Linux+Apache2+mysql+Php).

2.1.1.1 Apache2: It is a HTTP web server application in the internet world (World wide web) and is found to be secure and comprehensive in terms of features it provides. It allows virtual hosting i.e, Apache installation serves simultaneously, many different websites (Figure 7 & 8).

2.1.1.2 MYSQL: (Structure query language) It is a database management system which is relational database that stores data in tables for easy access and data retrieval. Its found to be an open source whose software can be used or modified by ayone. It is very fast, reliable and works in server/client systems.

2.1.1.3 PHP: (Hyper text pre processor) It is a server sided scripting language. This is essential for web development as the code is executed on the server, generating html that is sent to the clients later. All the html files can be processed with PHP. It is simple and has advanced features. This is used for web designing of database driven websites.

HTML Files: (Hyper Text Markup Language) which is a computer language that directs the formatting of text pages which will be displayed as webpages. This helps to use fonts, colors, create headlines, and graphics. It is a predominant language of world wide web. It is a coding language used to create our own website.

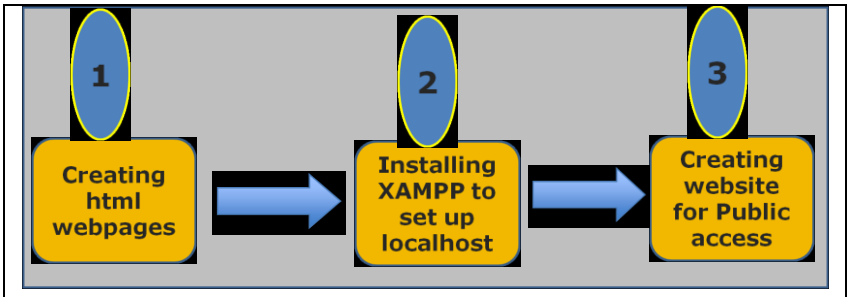


Figure 7. Steps involved in implementation of webserver.

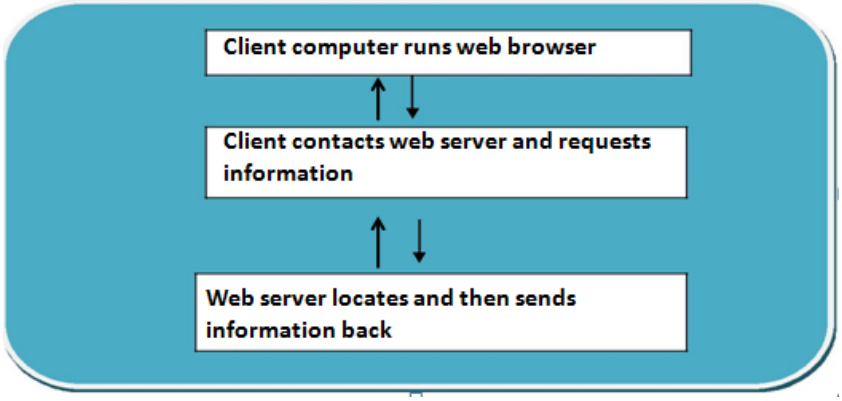


Figure 8. Client-server working protocol.

2.2 DEVELOPMENT OF EK3D, AN *E. COLI* K ANTIGEN 3D STRUCTURAL DATABASE

2.2.1 Collection of data

The capsular antigen (K antigen) of *E. coli* has been extensively studied with >80 antigen structures known so far (Orskov et al., 1977). From a preliminary literature survey, we found that though the carbohydrate composition and connectivity of these sugars had been determined, there was a marked lack of structural information. As a first step towards the development of *E. coli* K antigen, these antigen structures were collected from scientific literature and other web resources. A list of the K antigens and the corresponding references is shown in Table 4.

2.2.2 Generation of *E. coli* K antigen repeating units & their energy minimization

The *E. coli* K antigen repeating units were generated using GLYCAM, SWEETDB and manual pymol builder. The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC, (Loss et al., 2002).

2.2.3 CHARMM force field

CHARMM (Chemistry at HARvard Macromolecular Mechanics) is a program widely used for macromolecular mechanics and dynamics with potential energy functions for carbohydrates, proteins, nucleic acids, and lipids. The molecular system potential energy at a time t_1 can be expressed as a function of atomic positions using the familiar energy terms in Equation ……(1)

$V(r_1, \ldots, r_N) - \sum_{i<j} Vbonds(ri\ j) + \sum_{i<j<k} Vangles(\theta i\ jk) + \sum_{i<j<k<h} Vdihedral(\varphi i\ jkh) +$

$\sum_{i<j} Vcoul.(\ r_{ij}) + \sum_{i<j} V_{LJ}\ (\ r_{ij})$.....(1)

$Vbonds(r_{ij}) - \frac{1}{2}\ k^b_{\ i\ j}(r_{ij} \_\ r^0_{\ ij})^2$ ……(2)

$Vangles(\theta_{ijk}) - \frac{1}{2}k^{\theta}_{ijk}(\theta_{ijk}-\theta^0_{ijk})^2$ ……(3)

$Vdihedral(\varphi_{ijkl}) - \frac{1}{2}k_{\varphi}(1+ \cos(n_{\varphi}+\gamma))$ ……..(4)

$k^b_{\ i\ j}$ and $r^0_{\ i\ j}$ indicate the bond-stretching constant and the equilibrium distance from eq-2

k^{θ}_{ijk} and θ^0_{ijk} indicate the angle-bending constant and the equilibriumangle from eq-3

k_φ , n_φ and γ are the torsion constant, multiplicity and phase angle, respectively.

Where as Vbonds, Vangles, Vdihedrals refer to the potential energy with respect to bond stretching, angle bending, and dihedral angle rotations. VCoul and VLJ refer to the pairwise electrostatic interaction and to the Lennard-Jones (LJ) repulsion–dispersion potential terms, respectively. Classical force fields are defined by the functional form of these components and by the set of parameters that each term requires.

2.2.3.1 Parameterization of unusual sugars

Protein and nucleic acids are made from linear assembly of residues whereas because of the conformational flexibility of glycans, development of carbohydrate forcefield is challenging. Computational methods based on empirical force fields can contribute to the better understanding of structure and dynamics of carbohydrates. The presence of chiral centres in monosaccharides and branched nature of glycans makes the parameterization and the validation a difficult task. Considering two hexopyranose monosaccharides nearly 20 distict dissacharides can be formed where as only one dipeptide can be generated from the two aminoacids.

CHARMM additive all atom empericial force field for pyranoses and furanoses is built based on existing CHARMM parameters (Hatcher et al., 2009).

CHARMM CARBOHYDRATE PARAMETER FILE :

This describes parameters in the CHARMM empirical energy function. The CHARMM parameter file contains the information necessary to calculate potential energy. Parameterization was carried out for few unusual sugars that are not available in the CHARMM force field such as KDO, Quip-N4-Malate & hexulosonic acid. Following IUPAC nomenclature, graphical representation of the repeating units are developed

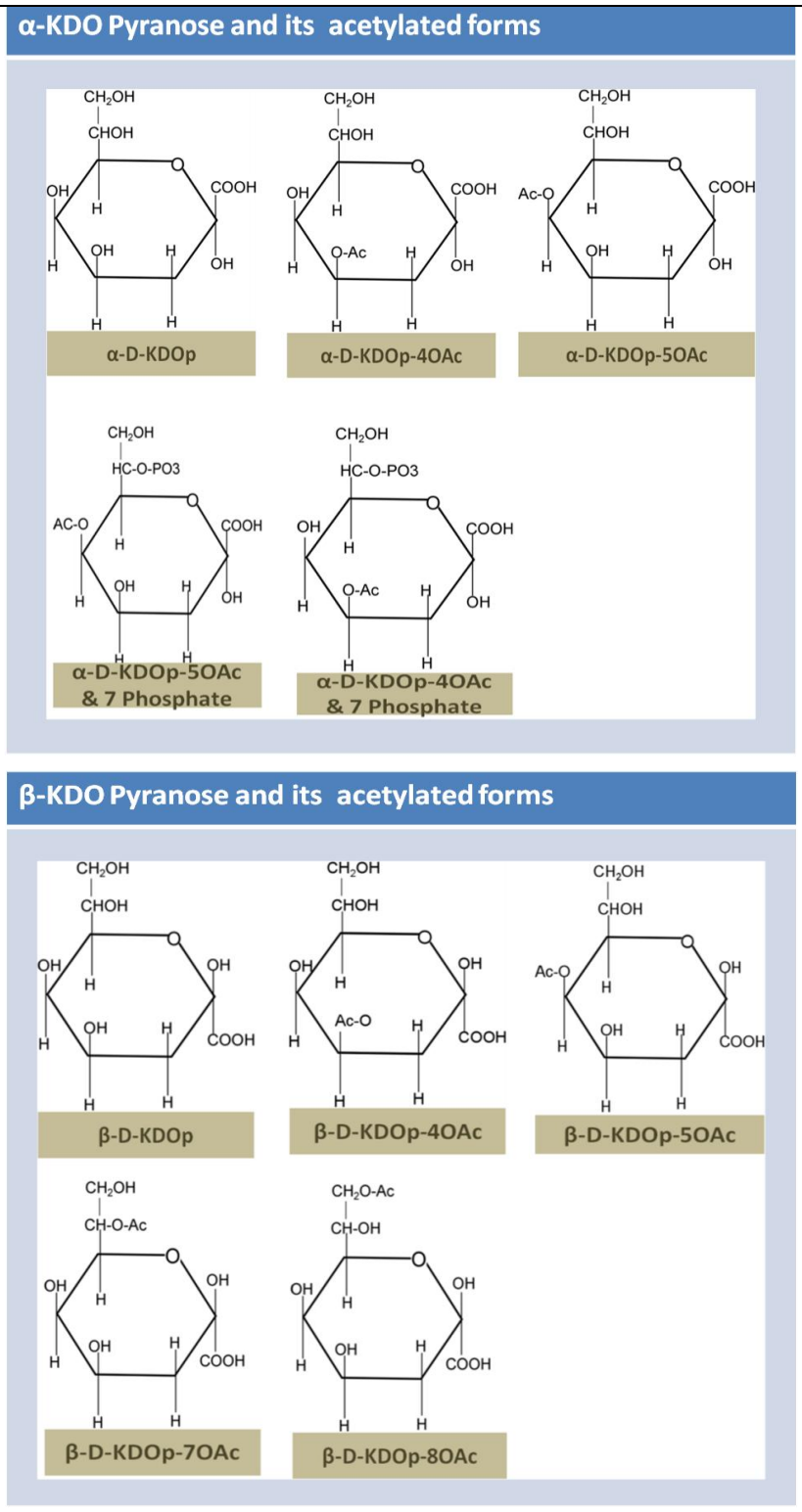


Figure 9 &10. Schematic representation of KDO pyranose forms and its derivatives.

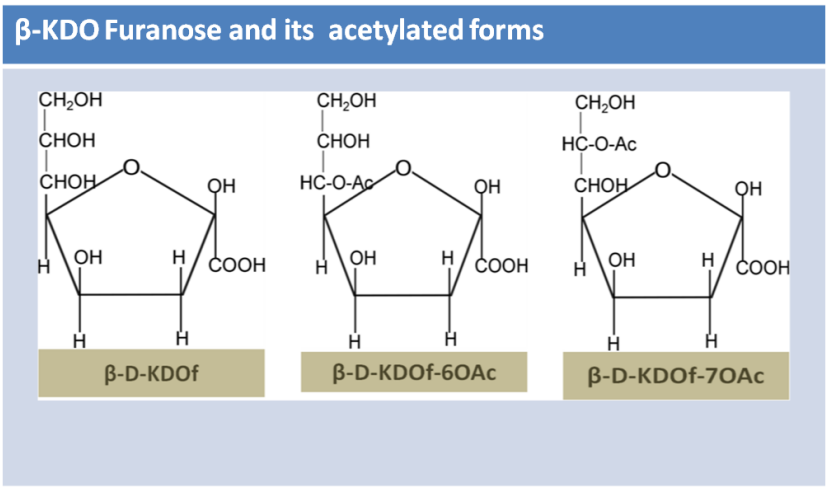


Figure 11. Schematic representation of KDO furanose forms.

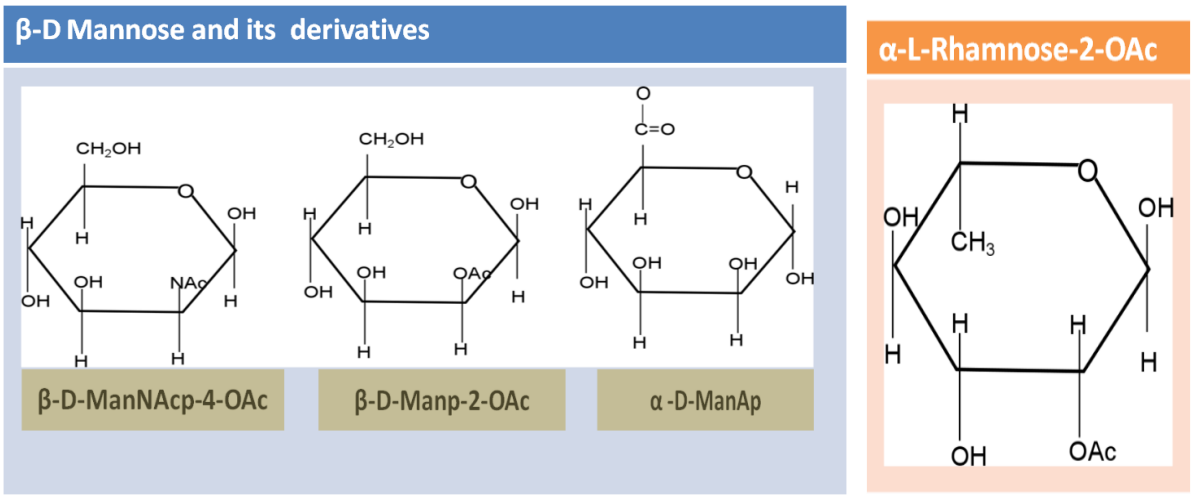


Figure 12. Schematic representation of mannose and its derivatives on left side and rhamnose derivative on the right side.

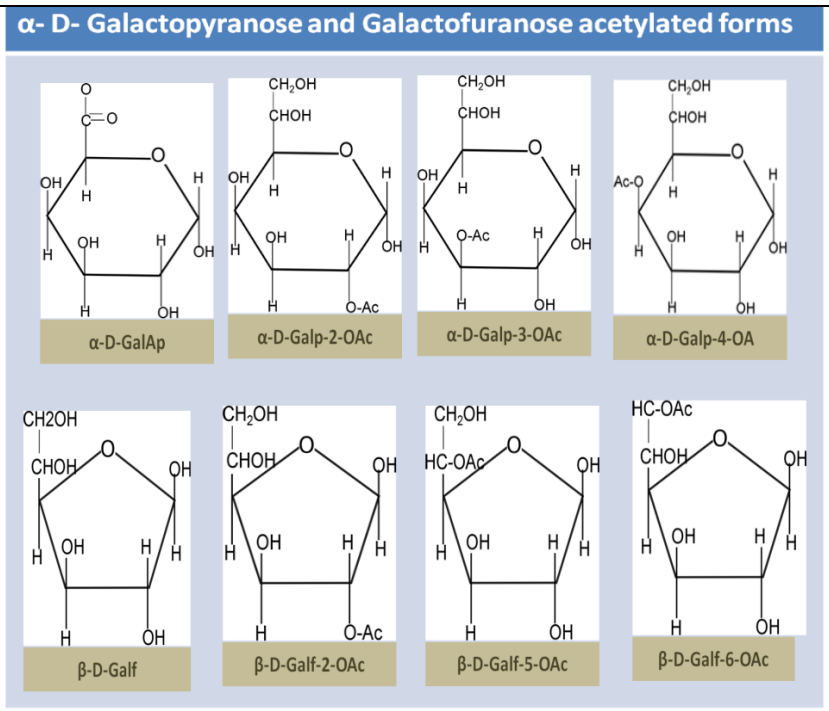


Figure 13. Schematic representation of galactose and its derivatives.

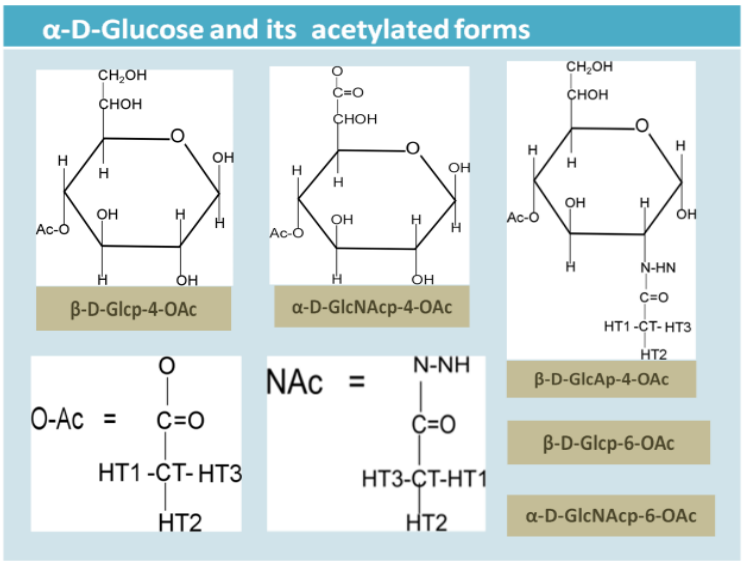


Figure 14. Schematic representation of glucose and its derivatives.

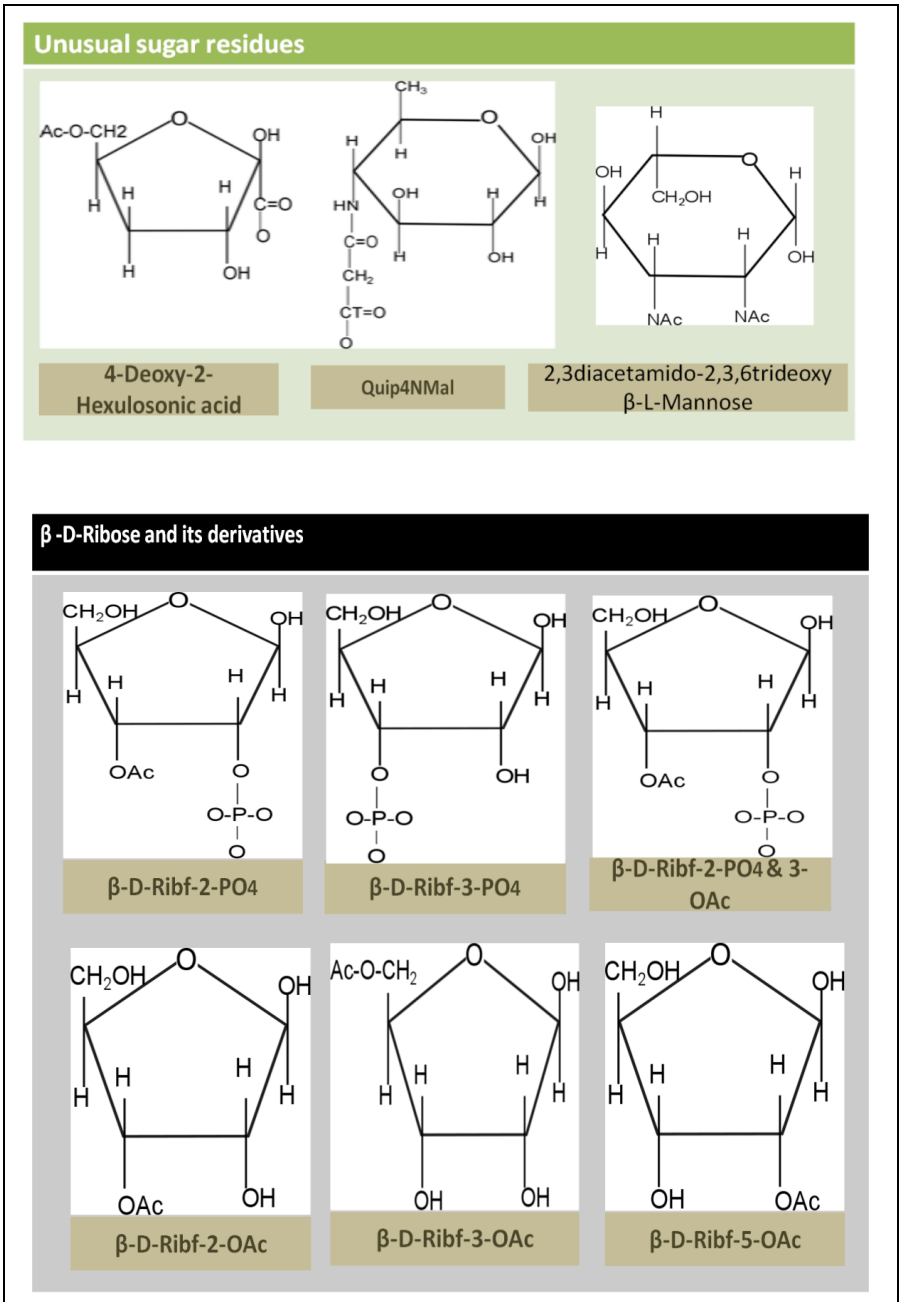


Figure 15. Schematic representation of Ribose and its derivatives.

2.2.3.2 Minimization of *E. coli* K antigen repeating unit

Minimization enables the reduction of short contacts and refinement of structure with an output file having no steric hindrance. Steepest descent (SD) is a gentle minimizer of molecules. Initially the PDB files should be given as an input for the CHARMM to get the structure read.

Steepest descent : SD is one of the energy minimization method employed to locate potential energy of atoms. The distorted geometries of the atoms can be repaired by few cycles of the steepest descent method (1500 steps were used). It quickly removes bad contacts and improves theconformation.
(http://www.uoxray.uoregon.edu/local/manuals/biosym/discovery/General/Minimization/Min_Algo.html)

ABNR / Newton-Raphson method : 1500 steps of minimization was carriedout after SD. (http://www.charmmtutorial.org/index.php/Minimization).

Approximate Quadratic energy function is given by Taylor series as :

$$f(x)=f(P)-bx+1/2Ax2$$

P= Current point

x= Arbitory point on the energy surface

b= Gradient at P

A= Hessian matrix at P

2.2.3.3 Representation of data in the database

Glycan representation

Compared to proteins, glycan representation is found to be more complex due to branching. Following IUPAC nomenclature graphical representation of repeating unit was developed. The representation includes monosaccharide anomericity, linkages, substitutions, and modifications of sugar residues.

Graphical Representations

Though several graphical representations are available in different glyco-focused databases the present format uses the cartoon representation for the symbols of monosaccharide names.

(i) Pictorial representation

According to IUPAC, for symbol representation of glycan structures, each sugar type should have the same symbol shape (Varki et al., 2009).

-Isomers must be differentiated by colour change or shading.

-Derivatives of sugars must be coloured with similar colours or with their shades

-Representation of sugars with similar shapes with different orientation should be avoided

- Linkage information and conformation of sugar is represented in the form of text

- The sugar symbols are attached to form the linear or branched glycan structure (Figure 16).

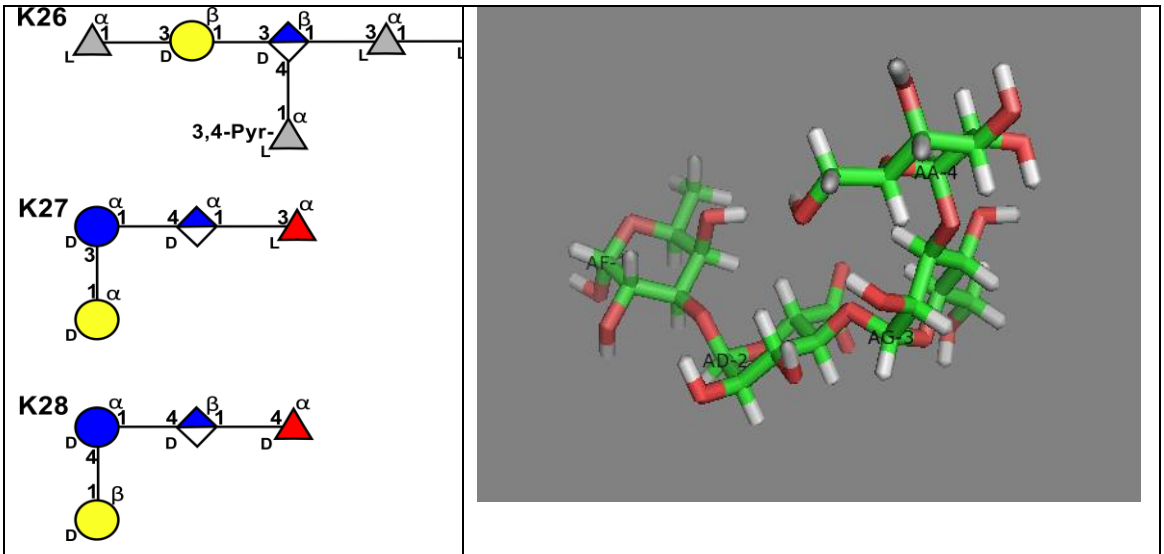


Figure 16. Left side view, pictorial representation of Group-1 capsules K26, K27, K28 and right side view, 3D pymol view of K antigen structure.

(ii) Naming scheme developed for EK3D

A four letter code was developed for naming the sugar residues based on IUPAC nomenclature. The first letter represents the alpha or beta configuration of the sugar. Second position is assigned to the sugar name (i.e, Glucose-G, Galactose-A, Mannose-M, Fucose-F, Rhamnose-H, KDO- K, Fructose-C, etc). Third position represents the acetylation/ phosphorylation of residues. Alphabets from A to H denote positions of O-acetylation from 1-8. Alphabets from I-P are assigned for N-acetylation positions from 1-8. Phosphorylation positions 1-5 are given as Q-U. X and Y alphabets represent mixture of O-acetylation, N-acetylation and phosphorylation. Combination of phosphorylation and acetylation are given numbers from 1-10 according to the existing glycan structures. Aminoacids like Serine and threonine are given as SER and THR to satisfy the naming nomenclature of proteins. The fourth position in naming is given to the presence of pyranose /furanose form of sugar.

A	G	B	0
A	Glucose	O-Acetylation at 2^{nd} position	By default pyranose

Jmol representation

The 3D structures of molecules can be viewed through Jmol. It is an interactive web browser applet for animated display of molecules. The PDB files of the antigen have to be placed in the specific directory for molecular visualization. The downloaded structure files can be viewed via Jmol application.

2.2.4 Multimer generation

The polymeric structure of *E. coli* group capsules can be obtained by translational method. Twist and atoms involved in linkage are noted and deletion of water may be user defined. Multimeric repeat units of varying length can be given as input.

Table 6. (A) Four letter code developed for naming residues for minimization of glycan structures (B) Sugar residues (one letter code) is shown position wise.

(A)

1st Position Anomeric Conformation	2nd position Residue Name	3rd position OAc/Nac/P	4th position Furanose/Pyranose

(B)

S.No	Residue Name	Letter code
1	Glucose	G
2	Galactose	A
3	Mannose	M
4	Rhamnose	H
5	Fucose	F
6	Quinovose	Q
7	Glucouronic acid	D
8	Galactouronic acid	E
9	Neuraminic acid	N
10	Ribitol	B
11	Glycerol	Y
12	KDO	K
13	Pyruvate	P
14	Ribofuranose	R
15	Fructose	C
16	Hex-2-ulosonic acid	S
17	Serine	SER
18	Threonine	THR

S.No	Position	Residue Name	Letter code
19	1-8	O-Acetylation	A-H
20	1-8	N-Acetylation	I-P
21	1-7	Phosphorylation	Q-W
22		1PO4 + 2NAc + 4OAc	X
23		1PO4 + 2NAc + 6OAc	Y
24		1PO4 + 2OAc	1
25		1PO4 + 3OAc	2
26		1PO4 + 4OAc	3
27		1PO4 + 6OAc	4
28		2PO4 + 3OAc	5
29		7PO4 + 4OAc	6
30		7PO4 + 5OAc	7
31		2NAc + 3OAc	8
32		2NAc + 4OAc	9
33		2NAc + 6OAc	10

Table 7. List of unavailable residues developed for K antigen structures minimization.

NEW RESIDUES	TOPOLOGY
RESIDUE NAME	**DESCRIPTION**
BGD	Beta-D-Glucose + 4 Oacetylation
BDD	Beta-D-GlcpA + 4 Oacetylation
AGJ	Alpha-D-GlcpNAc
AG9	Alpha-D-GlcNAc + 6 OAcetylation
AAD	Alpha-D-GalpA
AAB	Alpha-D-Galp + 2 Oacetylation
AAC	Alpha-D-Galp + 3 OAcetylation
AAD	Alpha-D-Galp + 4 Oacetylation
AA0F	Alpha-D-Galf
AABF	Alpha-D-Galf + 2 Oacetylation
BMJ	Beta-D-ManpNAc
BMB	Beta-D-Manp + 2 OAcetylation
AMJ	Alpha-D-ManpNAc
AHB	Alpha-L-Rhap + 2 Oacetylation
AF	Alpha-D-Fucp
BF	Beta-D-Fucp
BFK	Beta-D-Fucp3NAc
AFB	Alpha-L-Fucp + 2 Oacetylation
AFC	Alpha-L-Fucp + 3 Oacetylation
AK	Alpha-D-Kdop
AKD	Alpha-D-Kdop + 4 Oacetylation
AKE	Alpha-D-Kdop + 5 Oacetylation
AK6	Alpha-D-Kdop + 4 OAcetylation + 7 Phosphate
AK7	Alpha-D-Kdop + 5 OAcetylation + 7 Phosphate
BK0F	Beta-D-Kdof
	4-Deoxy 2-Hexulosonic acid + 6 Oacetylation
BQ	Beta-D-Quip4NMal (4,6-Dideoxy-4-malonyl amino glucose)
	Beta-L-Sugarp (2,3-Diacetamide-,3,6-trideoxy-β-L-mannopyranose
BRBF	Beta-D-Ribf + 2 Oacetylation
BRCF	Beta-D-Ribf + 3 Oacetylation
BREF	Beta-D-Ribf + 5 Oacetylation
BRR	Beta-D-Ribf + 2 Phosphate
BRS	Beta-D-Ribf + 3 Phosphate
BR5F	Beta-D-Ribf + 2 Phosphate + 3 OAcetylation

Table 8. Overview of patches used in minimizing K antigen structures

List of Patches developed for minimizing glycan structures

NEW PATCHES	DESCRIPTION
PYRUVATE	**PATCHES**
23 PS	Pyruvate (S Configuration)
23 PR	Pyruvate (R Configuration)
43 PR	Pyruvate (R Configuration)
43PS	Pyruvate (S Configuration)
46PR	Pyruvate (R Configuration)
46PS	Pyruvate (S Configuration)
KDO	**PATCHES**
K1-4	Furanose-KDOpyranose
K1-5	Pyranose-KDOpyranose
K1-6	Furanose-KDOfuranose
K1-7	Furanose-KDOpyranose
K1-8	Furanose-KDOfuranose
KG21	KDOpyranose-Glycerol
GLYCEROL	**PATCHES**
1GP	Glycerol-1-Phosphate
1-2PG	Pyranose-Glycerol
1-2FG	Furanose-Glycerol
1-4GPP	Glycerol-Phosphate-Pyranose
1-5GPF	Glycerol-Phosphate-Furanose
PHOSPHATE	**PATCHES**
1PP	Pyranose-1-Phosphate
PYR /FURANOSE	**PATCHES**
1-2PF	Pyranose-Furanose
1-4FP	Furanose-Pyranose
2-3FP	Furanose-Pyranose
2-2FP	Furanose-Pyranose
1-3FF	Furanose-Furanose
PP29	Pyranose-Pyranose

List of references for K antigen structures

K1-
https://books.google.co.in/books?id=s6gbXT15CeYC&pg=PA108&lpg=PA108&dq=Weintraub,+A.+(2003)+Carbohydr.+Res.,+338,+2539+2547&source=bl&ots=ED1WD34Bjp&sig=ri8omlOq8hJM2PhCKKz7nFAlkwY&hl=en&sa=X&ved=0CB0Q6AEwAGoVChMIjvvJ1PDZxgIVIiCmCh0CKAyF#v=onepage&q=E.coli&f=false page 353

K1, K2a, K5, K12/K82, K13, K27, K29, K30, K40, K92, K100- - (Whitfield, 2006)

K2ab, K4, K7 & K56, K8, K11, K13, K14, K16, K18, K19, K20, K22, K51, K74, K93, -
http://link.springer.com/chapter/10.1007/978-3-642-74694-9_2#page-1

K3- (Dengler et al., 1988)

K6-(Jennings et al., 1982)- Request

K9-(Dutton et al., 1987)

K10-(Sieberth et al., 1993)

K12, K82, K23, K52, K57- (Ovodov, 2006)

K15-(Jann and Jann, 1987)

K24-(Lenter et al., 1990)

K25-(Anderson et al., 1994)

K26- (Beynon and Dutton, 1990)

K28-(Altman and Dutton, 1985)

K31-(Dutton et al., 1990)

K32-(Annison et al., 1987)

K33- (Leslie et al., 1998)

K34-(Dutton and Kuma-Mintah, 1987)

K35- (Hackland and Parolis, 1992)

K36- (Parolis et al., 1988)

K37-(Anderson et al., 1987)

K38- (Hackland et al., 1991)

K39- (Parolis et al., 1989)

K41- (Leslie et al., 1997)

K42-(Niemann et al., 1978)

K43-(Choy et al., 1995)

K44-(Dutton et al., 1988)

K46-(Dutton et al., 1992)

K47-(de Bruin et al., 1992b)

K48-(Whittaker et al., 1994a)

K49- (Beynon et al., 1990)

https://books.google.co.in/books?id=dd5xavr4bhmc&pg=pa215&dq=k1+structure+e.coli&hl=en&sa=x&ei=jfgzvym6adedugtn8zqidg&ved=0ccyq6aewag#v=onepage&q=k21%20structure%20e.coli&f=false page 216

K50- (Grue et al., 1994b)

K53—(Bax et al., 1988)

K54-(Hofmann et al., 1985)

K55-(Anderson and Parolis, 1989)

K57-(Vann et al., 1981, Ovodov, 2006)

K74-(Ahrens et al., 1988)

K83-(Whittaker et al., 1994b)

K84-(Whittaker et al., 1994c)

K87-(Parolis et al., 1990)

K93-(Bax et al., 1988)

K95-(Orskov and Orskov, 1976)

K96- (Jann et al., 1994)

K97-(Orskov and Orskov, 1976)

K98-(Hahne et al., 1991)

K101-(Grue et al., 1993)

K102- (de Bruin et al., 1992a)

K103- (Grue et al., 1994a)

REFERENCES

AHRENS, R., JANN, B., JANN, K. & BRADE, H. 1988. Structure of the K74 antigen from Escherichia coli O44:K74:H18, a capsular polysaccharide containing furanosidic beta-KDO residues. *Carbohydr Res,* 179**,** 223-31.

ALTMAN, E. & DUTTON, G. G. 1985. Chemical and structural analysis of the capsular polysaccharide from Escherichia coli O9:K28(A):H- (K28 antigen). *Carbohydr Res,* 138**,** 293-303.

ANDERSON, A. N. & PAROLIS, H. 1989. Investigation of the structure of the capsular polysaccharide of Escherichia coli K55 using Klebsiella bacteriophage phi 5. *Carbohydr Res,* 188**,** 157-68.

ANDERSON, A. N., PAROLIS, H. & PAROLIS, L. A. 1987. Structural investigation of the capsular polysaccharide from Escherichia coli O9:K37 (A 84a). *Carbohydr Res,* 163**,** 81-90.

ANDERSON, A. N., PAROLIS, L. A. & PAROLIS, H. 1994. Structural determination of the capsular antigen of Escherichia coli O8:K25:H9. *Carbohydr Res,* 265**,** 41-7.

ANNISON, G., DUTTON, G. G. & ALTMAN, E. 1987. Structure of the capsular polysaccharide of Escherichia coli O9:K32(A):H19. *Carbohydr Res,* 168**,** 89-102.

BAX, A., SUMMERS, M. F., EGAN, W., GUIRGIS, N., SCHNEERSON, R., ROBBINS, J. B., ORSKOV, F., ORSKOV, I. & VANN, W. F. 1988. Structural studies of the Escherichia coli K93 and K53 capsular polysaccharides. *Carbohydr Res,* 173**,** 53-64.

BEYNON, L. M. & DUTTON, G. G. 1990. The structure of Escherichia coli K26 antigen. *Carbohydr Res,* 200**,** 457-68.

BEYNON, L. M., DUTTON, G. G. & RICHARDS, J. C. 1990. Structure of the amino acid-containing capsular polysaccharide from Escherichia coli O8:K49:H21. *Carbohydr Res,* 205**,** 347-59.

CHOY, Y. M., DUTTON, G. G., LESLIE, M. R., PAROLIS, H. & PAROLIS, L. A. 1995. The structure of the capsular polysaccharide of Escherichia coli O8:K43:H11. *Carbohydr Res,* 269**,** 295-302.

DE BRUIN, A. H., PAROLIS, H. & PAROLIS, L. A. 1992a. The capsular antigen of Escherichia coli serotype O8:K102:H. *Carbohydr Res,* 235**,** 199-209.

DE BRUIN, A. H., PAROLIS, H. & PAROLIS, L. A. 1992b. Structural elucidation of the capsular polysaccharide of E. coli serotype K47. *Carbohydr Res,* 233**,** 195-204.

DENGLER, T., HIMMELSPACH, K., JANN, B. & JANN, K. 1988. Structure of the capsular K3 antigen of Escherichia coli 04:K3:H4, a polysaccharide containing a 4-deoxy-2-hexulosonic acid. *Carbohydr Res,* 178**,** 191-201.

DUTTON, G. G., KARUNARATNE, D. N. & LIM, A. V. 1988. Escherichia coli serotype K44: an acidic capsular polysaccharide containing two 2-acetamido-2-deoxyhexoses. *Carbohydr Res,* 183**,** 111-22.

DUTTON, G. G. & KUMA-MINTAH, A. 1987. Structure of Escherichia coli capsular antigen K34. *Carbohydr Res,* 169**,** 213-20.

DUTTON, G. G., KUMA-MINTAH, A., NG, S. K., PAROLIS, H., DELL, A. & REASON, A. 1992. Determination of the structure of the capsular antigen of Escherichia coli O8:K46:H30, using FABMS and 2D-NMR spectroscopy. *Carbohydr Res,* 231**,** 39-50.

DUTTON, G. G., KUMA-MINTAH, A. & PAROLIS, H. 1990. The structure of Escherichia coli K31 antigen. *Carbohydr Res,* 197**,** 171-80.

DUTTON, G. G., PAROLIS, H. & PAROLIS, L. A. 1987. The structure of the neuraminic acid-containing capsular polysaccharide of Escherichia coli serotype K9. *Carbohydr Res,* 170**,** 193-206.

GRUE, M. R., PAROLIS, H. & PAROLIS, L. A. 1993. Structural elucidation of the capsular polysaccharide of Escherichia coli serotype K101 by high resolution NMR spectroscopy. *Carbohydr Res,* 246**,** 283-90.

GRUE, M. R., PAROLIS, H. & PAROLIS, L. A. 1994a. Structural investigation of the capsular polysaccharide of Escherichia coli O101 : K103 : H- using bacteriophage degradation and NMR spectroscopy. *Carbohydr Res,* 262**,** 311-22.

GRUE, M. R., PAROLIS, H. & PAROLIS, L. A. 1994b. Structure of the capsular antigen of Escherichia coli O8:K50:H. *Carbohydr Res,* 258**,** 233-41.

HACKLAND, P. L. & PAROLIS, H. 1992. The structure of the capsular polysaccharide of Escherichia coli O9:K35:H. *Carbohydr Res,* 235**,** 211-20.

HACKLAND, P. L., PAROLIS, H. & PAROLIS, L. A. 1991. Escherichia coli O9:K38 capsular antigen: another ribofuranose-containing glycan. *Carbohydr Res,* 219**,** 193-201.

HAHNE, M., JANN, B. & JANN, K. 1991. Structure of the capsular polysaccharide (K98 antigen) of E. coli O7:K98:H6. *Carbohydr Res,* 222**,** 245-53.

HATCHER, E., GUVENCH, O. & MACKERELL, A. D. 2009. CHARMM additive all-atom force field for aldopentofuranoses, methyl-aldopentofuranosides, and fructofuranose. *J Phys Chem B,* 113**,** 12466-76.

HOFMANN, P., JANN, B. & JANN, K. 1985. Structure of the amino acid-containing capsular polysaccharide (K54 antigen) from Escherichia coli O6:K54:H10. *Carbohydr Res,* 139**,** 261-71.

JANN, B., KOCHANOWSKI, H. & JANN, K. 1994. Structure of the capsular K96 polysaccharide (K96 antigen) from Escherichia coli O77:K96:H- and comparison with the capsular K54 polysaccharide (K54 antigen) from Escherichia coli O6:K54:H10. *Carbohydr Res,* 253**,** 323-7.

JANN, K. & JANN, B. 1987. Polysaccharide antigens of Escherichia coli. *Rev Infect Dis,* 9 Suppl 5**,** S517-26.

JENNINGS, H. J., ROSELL, K. G. & JOHNSON, K. G. 1982. Structure of the 3-deoxy-D-manno-octulosonic acid-containing polysaccharide (K6 antigen) from Escherichia coli LP 1092. *Carbohydr Res,* 105**,** 45-56.

LENTER, M. C., JANN, B. & JANN, K. 1990. Structure of the K24 antigen of E. coli O83:K24:H, a polymer that consists of alpha-Kdop and glycerol phosphate. *Carbohydr Res,* 208**,** 139-44.

LESLIE, M. R., PAROLIS, H. & PAROLIS, L. A. 1997. Structural analysis of the capsular antigen of Escherichia coli O8:K41:H11. *Carbohydr Res,* 299**,** 197-202.

LESLIE, M. R., PAROLIS, H., PAROLIS, L. A. & PETERSEN, B. O. 1998. The capsular antigen of Escherichia coli O9:K33:H-: a polysaccharide containing both pyruvate and O-acetyl groups. *Carbohydr Res,* 309**,** 95-101.

LOSS, A., BUNSMANN, P., BOHNE, A., LOSS, A., SCHWARZER, E., LANG, E. & VON DER LIETH, C. W. 2002. SWEET-DB: an attempt to create annotated data collections for carbohydrates. *Nucleic Acids Res,* 30**,** 405-8.

MERRITT, J. H., OLLIS, A. A., FISHER, A. C. & DELISA, M. P. 2013. Glycans-by-design: engineering bacteria for the biosynthesis of complex glycans and glycoconjugates. *Biotechnol Bioeng,* 110**,** 1550-64.

NIEMANN, H., CHAKRABORTY, A. K., FRIEBOLIN, H. & STIRM, S. 1978. Primary structure of the Escherichia coli serotype K42 capsular polysaccharide and its serological identity with the Klebsiella K63 polysaccharide. *J Bacteriol,* 133**,** 390-1.

ORSKOV, I. & ORSKOV, F. 1976. Five new Escherichia coli K antigens, K95, K96, K97, K98 and K100. *Acta Pathol Microbiol Scand B,* 84B**,** 321-5.

ORSKOV, I., ORSKOV, F., JANN, B. & JANN, K. 1977. Serology, chemistry, and genetics of O and K antigens of Escherichia coli. *Bacteriol Rev,* 41**,** 667-710.

OVODOV, Y. S. 2006. Bacterial capsular antigens. Structural patterns of capsular antigens. *Biochemistry (Mosc),* 71**,** 937-54.

PAROLIS, H., PAROLIS, L. A. & STANLEY, S. M. 1988. The use of bacteriophage-mediated depolymerisation in the structural investigation of the capsular polysaccharide from Escherichia coli serotype K36. *Carbohydr Res,* 175**,** 77-83.

PAROLIS, H., PAROLIS, L. A., STANLEY, S. M. & DUTTON, G. G. 1990. The structure of the capsular antigen from Escherichia coli O8:K87:H19. *Carbohydr Res,* 205**,** 361-70.

PAROLIS, H., PAROLIS, L. A. & VENTER, R. D. 1989. Escherichia coli serotype-39 capsular polysaccharide: primary structure and depolymerisation by a bacteriophage-associated glycanase. *Carbohydr Res,* 185**,** 225-32.

SIEBERTH, V., JANN, B. & JANN, K. 1993. Structure of the K10 capsular antigen from Escherichia coli O11:K10:H10, a polysaccharide containing 4,6-dideoxy-4-malonylamino-D-glucose. *Carbohydr Res,* 246**,** 219-28.

VANN, W. F., SCHMIDT, M. A., JANN, B. & JANN, K. 1981. The structure of the capsular polysaccharide (K5 antigen) of urinary-tract-infective Escherichia coli 010:K5:H4. A polymer similar to desulfo-heparin. *Eur J Biochem,* 116**,** 359-64.

VARKI, A., CUMMINGS, R. D., ESKO, J. D., FREEZE, H. H., STANLEY, P., MARTH, J. D., BERTOZZI, C. R., HART, G. W. & ETZLER, M. E. 2009. Symbol nomenclature for glycan representation. *Proteomics,* 9**,** 5398-9.

WHITFIELD, C. 2006. Biosynthesis and assembly of capsular polysaccharides in Escherichia coli. *Annu Rev Biochem,* 75**,** 39-68.

WHITTAKER, D. V., PAROLIS, L. A. & PAROLIS, H. 1994a. Escherichia coli K48 capsular polysaccharide: a glycan containing a novel diacetamido sugar. *Carbohydr Res,* 256**,** 289-301.

WHITTAKER, D. V., PAROLIS, L. A. & PAROLIS, H. 1994b. Structural elucidation of the capsular polysaccharide expressed by Escherichia coli O20:K83:H26 by high resolution NMR spectroscopy. *Carbohydr Res,* 253**,** 247-56.

WHITTAKER, D. V., PAROLIS, L. A. & PAROLIS, H. 1994c. Structural elucidation of the capsular polysaccharide produced by Escherichia coli O20 : K84 : H26. *Carbohydr Res,* 262**,** 323-34.

http://glycam.org/docs/articlesandciting/citing

http://www.uoxray.uoregon.edu/local/manuals/biosym/discovery/General/Minimization/Min_Algo.html

http://www.charmmtutorial.org/index.php/Minimization

http://glycomics.scripps.edu/CFGnomenclature.pdf

CHAPTER 3

RESULTS

We have developed EK3D as a comprehensive bank of *E. coli* K antigen structural data. Structures of 72 *E. coli* K antigens distributed across the 4 capsular groups have been provided. This is the first database with a specific focus on K antigen 3D structures. The main features of the database are described below.

3.1 Implementation of EK3D as a database

EK3D has emerged as a database management with necessary requirements serving as a bioinformatics tool providing the K antigen data of an important model organism like *E. coli*. The navigation bar offers different sections related to the *E. coli* K antigen study. Initially an introduction to capsular polysaccharide antigens is displayed in About K antigen section. The later section that is classification involves a brief discussion on the *E. coli* groups classification based on capsular biosynthesis and surface assembly. Links to different group biosynthesis mechanism, proteins involved are clearly depicted along with the pictorial representation. The group properties are listed in a table making it easy to understand. K antigen structures section offers structural representation of nearly more than 80 *E. coli* serotypes of different groups (Ovodov, 2006). The structures display type of sugars present, linkages & branching. To facilitate the interpretation of structure results, a file download page was developed. This will help users to download the required K antigen structure file and view the 3D model. Data on the available protein structural information like 2D, 3D and DSSP (secondary structural assignments) is provided in protein structures. The generation of polymeric CPS structures of varying lengths for modeling and molecular dynamic investigations has also been offered for users.

3.2 CREATING TOPOLOGY FILES FOR UNUSUAL SUGARS

KDO TOPOLOGY FILE (RTF FILE)

ATOM NAME ATOM TYPE CHARGE

```
GROU
ATOM C1  CC2O2   0.300  !
ATOM O11 OC2D2  -0.600 !
ATOM O12 OC2D2  -0.600 !
ATOM C2  CC3062  0.330  !
ATOM O2  OC311  -0.650  !
ATOM HO2 HCP1   0.420  !
ATOM C6  CC3163  0.110  !
ATOM H6  HCA1    0.090  !
ATOM O6  OC3C61 -0.400 !
GROU                    !
ATOM C3  CC3261 -0.180  !
ATOM H31 HCA2    0.090  !
ATOM H32 HCA2    0.090  !
GROU                    !
ATOM C4  CC3161  0.140  !
ATOM H4  HCA1    0.090  !
ATOM O4  OC311  -0.650  !
ATOM HO4 HCP1   0.420  !
GROU                    !
ATOM C5  CC3161  0.140  !
ATOM H5  HCA1    0.090  !
ATOM O5  OC311  -0.650  !
ATOM HO5 HCP1   0.420  !
GROU                    !
ATOM C7  CC312   0.14   !
ATOM H7  HCA1    0.09   !
ATOM O7  OC311  -0.65   !
ATOM HO7 HCP1   0.42   !
GROU                    !
ATOM C8  CC322   0.05   !
ATOM H81 HCA2    0.09   !
ATOM H82 HCA2    0.09   !
ATOM O8  OC311  -0.65   !
ATOM HO8 HCP1   0.42   !
!
BOND C1 O11    C1 O12    C1 C2    C2 O2
O2 HO2
BOND C2 C3     C3 H31    C3 H32   C3 C4
C4 H4
BOND C4 C5     C4 O4     O4 HO4   C5 H5
C5 O5
BOND O5 HO5    C5 C6     C6 H6    C6 C7
C7 H7
BOND C7 O7     C6 O6     O6 C2    O7 HO7
C7 C8
BOND C8 H81    C8 H82    C8 O8    O8 HO8
PATC FIRS NONE LAST NONE
```

STRUCTURE

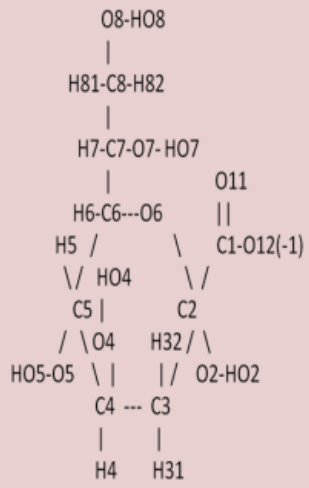


Figure 17. Newly developed KDO topology file. The atom numbers, atom types and their grouping pattern displaying the charges and bonds are clearly mentioned.

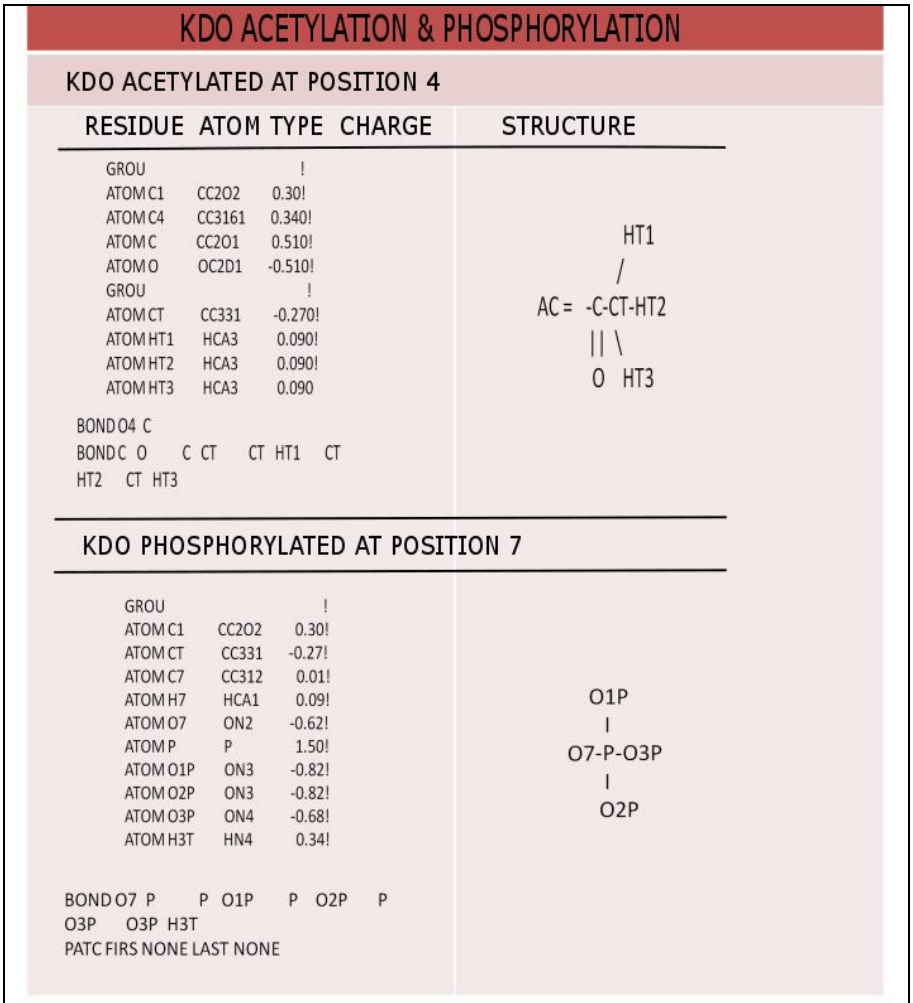
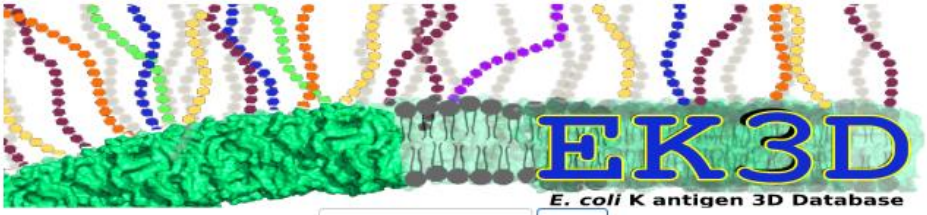


Figure 18. KDO acetylation and phosphorylation showing atom types and their charges.

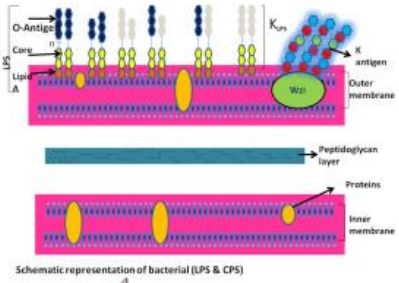
Screen images of EK3D



Enter search key Search

Welcome to EK3D

The database provides information about the group polysaccharide K antigens of E. coli, assembly mechanism, proteins involved in CPS surface assembly, generating cps structures of antigens, biophysical and mutagenesis structures



Schematic representation of bacterial (LPS & CPS)

Whats new !!!

K antigen structures of group 1, 2, 3 & 4

Related links

Carbohydrate databases

PDB (http://www.rcsb.org/pdb/home/home.do)

CATH (http://www.cathdb.info/)

Glycosciences (http://www.glycosciences.de/)

PolysacDB (http://crdd.osdd.net/raghava/polysacdb/index.html)

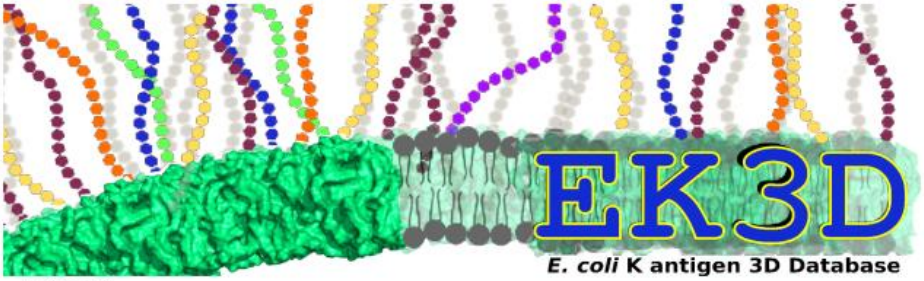
Bacterial Carbohydrate Structure Database (http://csdb.glycoscience.ru/bacterial/)

CFG-Glycan DB (http://www.functionalglycomics.org/glycomics/molecule/jsp/carbohydrate/carbMoleculeHome.jsp)

E. coli databases

ECODAB (http://nevyn.organ.su.se/ECODAB/)

EcoCyc (http://ecocyc.org/)



RESULTS:

WZA - Outermembrane lipo protein

()
Wza is an outer membrane protein involved in translocation of the capsular polysaccharide.It forms a channel and interacts with Wzc periplasmic domain.

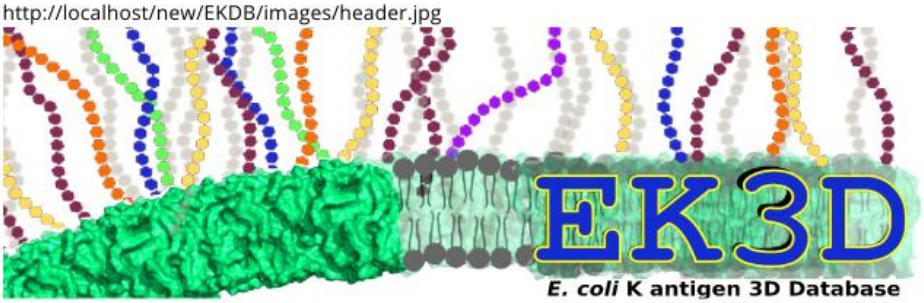
WZA-Exporter protein

http://www.rcsb.org/pdb/explore/jmol.do?structureId=2J58&bionumber=1 (http://www.rcsb.org/pdb/explore/jmol.do?structureId=2J58&bionumber=1)
The oligomeric structure of Wza is an octamer. The presence of Wza is essential for the assembly of the prototype group 1 capsule structure on the surface of E. coli serotype K30. Wza forms stable oligomeric "donut"-shaped complexes with a molecular mass of ~300 kDa, and a preliminary study of Wza, based on small two-dimensional crystalline areas, revealed ring like complexes with an average outer diameter of ~90-110 and a central stained region of ~20-35 in diameter. Wza is a tetramer of dimmers.

E.coli Group-1 biosynthesis

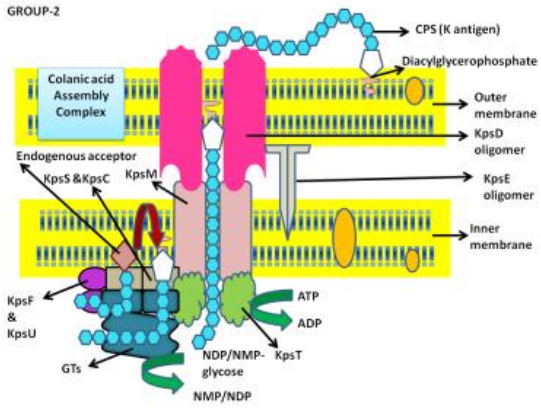
http://192.168.2.38/bharathi/index_files/Page772.html (http://192.168.2.38/bharathi/index_files/Page772.html)
Group-1 biosynthesis Assembly via the Wzx/Wzy-dependent pathway begins in the cytoplasm with the synthesis of lipid-linked polysaccharide repeat units. The first sugar galoctose-1-P from UDP galactose is added by an initiating glycosyl transferase (WbaP) to the lipid carrier molecule undecaprenyl phosphate (UndP) in the form of an activated nucleotide precursor at the inner leaflet of the inner membrane (IM), resulting in a pyrophosphate linkage with the carrier (undecaprenyl pyrophosphate,UndPP) ....more

http://localhost/new/EKDB/images/header.jpg



Group-2 biosynthesis

Schematic representation of Group-2 capsular polysaccharide biosynthesis and surface assembly



Group-2 Proteins

WbaP : Initial Glycosyl transferase

Wzz : Chain length dictator

WaaL : Polymerization terminator

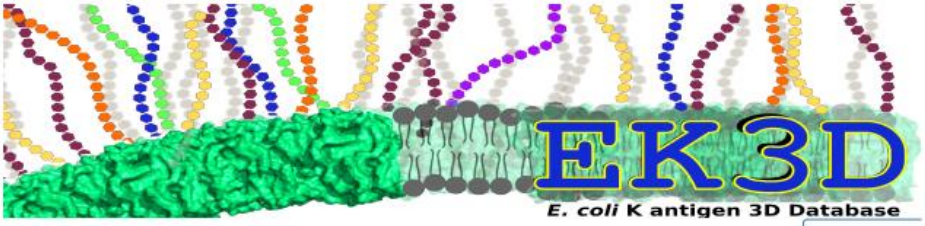
GTs : Glycosyl transferase

KpsMT : ABC transporters

KpsE : Adaptor protein

KpsD : Outer membrane translocon

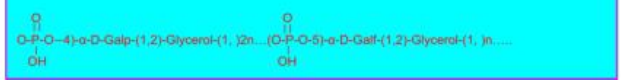
1. The group 2 and 3 biosynthesis occurs at kps loci near ser A and involves many conserved proteins. The genetic organization and regulatory features differ. The chromosomal loci for group-2 has 3 regions (Barrett B, Ebah L, Roberts IS, 2002). Region-1 encodes for proteins involved in export and assembly of capular polysaccharide. Region 2 is centrally located and serotype specific and codes for



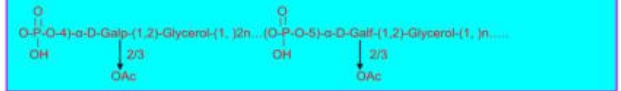
Group Polysaccharide Structures

K1

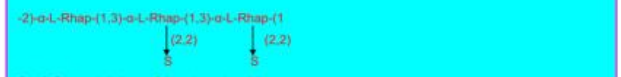
K2a



K2ab & K62



K3

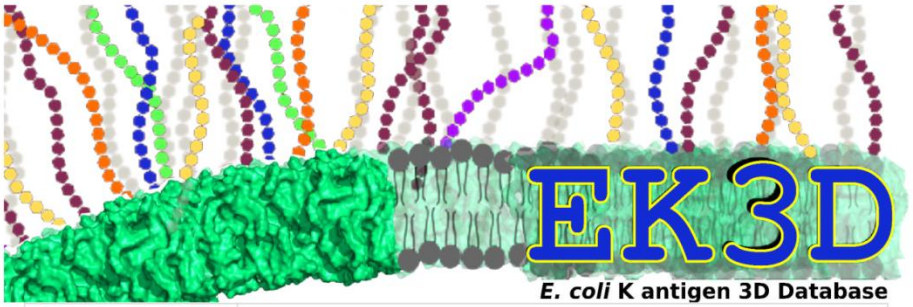


K4

K5

K6

K7 & K56



Name	Download
EK1	K antigen (files/k1.txt)
EK2ab	K antigen (files/k2ab.txt)
EK2a	K antigen (files/k2a.txt)
EK3	K antigen (files/k3.txt)
EK4	K antigen (files/k4.txt)
EK5	K antigen (files/k5.txt)
EK6	K antigen (files/k6)
EK7	K antigen (files/k7.txt)
EK8	K antigen (files/k8.txt)
EK9	K antigen (files/k9.txt)
EK10	K antigen (files/k10.txt)
EK11	K antigen (files/k11.txt)
EK12	K antigen (files/k12)
EK13b	K antigen (files/k13b.txt)
EK13a	K antigen (files/k13a)
EK14	K antigen (files/k14)
EK15	K antigen (files/k15)

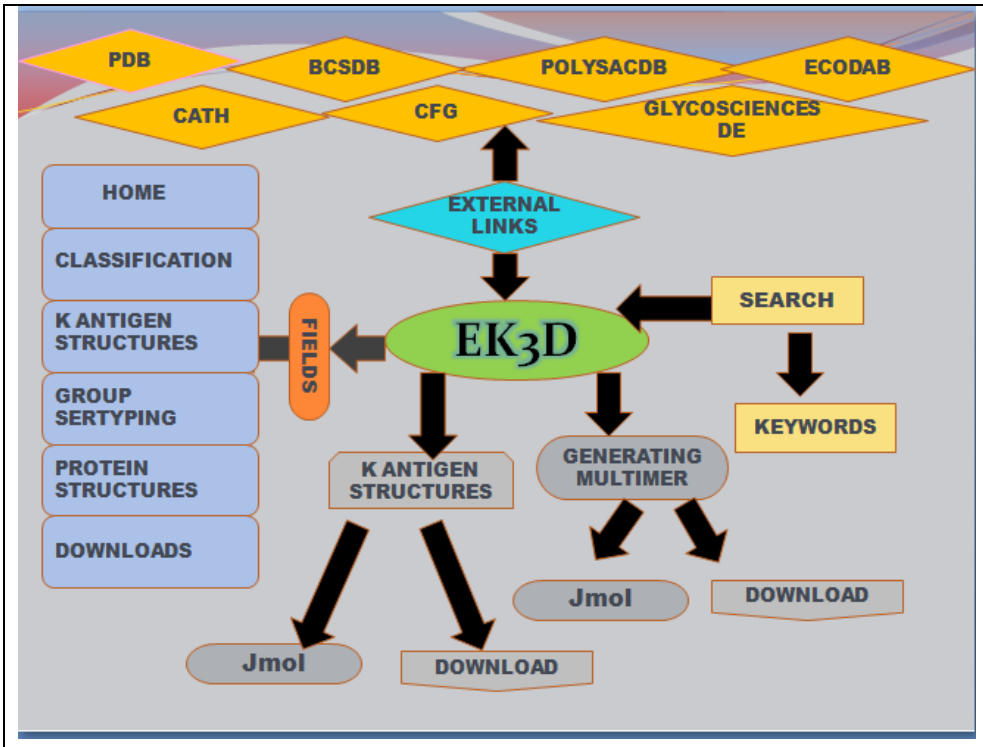


Figure 24. EK3D sitemap showing related links on the top, navigation bar on the left & other features of the database at the bottom.

3.2 Conclusion

Capsular polysaccharides are major virulence determinants of Gram-negative bacteria. They act as a physical barrier to host immune cell components and aid in immune evasion. The capsule has thus become an important antibacterial drug target. With the emergence of MDR bacterial strains, the need to develop new, efficient drugs has become even more significant. A large number of CPS structures have been discovered to date. For instance, in the well studied Gram-negative bacterium, *E. coli*, there are about 80 K antigen structures. The high antigenic variability of the K antigen thus poses a major challenge. Though a large amount of information is available about the CPS from different organisms, it has not been efficiently organized. A centralized database of this information would be extremely valuable for researchers. The database EK3D that we have developed is an important step in this direction. It provides comprehensive structural information about the Group 1-4 K antigens of *E. coli* and about the proteins involved in its biosynthesis. Thus, the database developed would facilitate the designing of new drugs for therapeutic use and modeling studies.