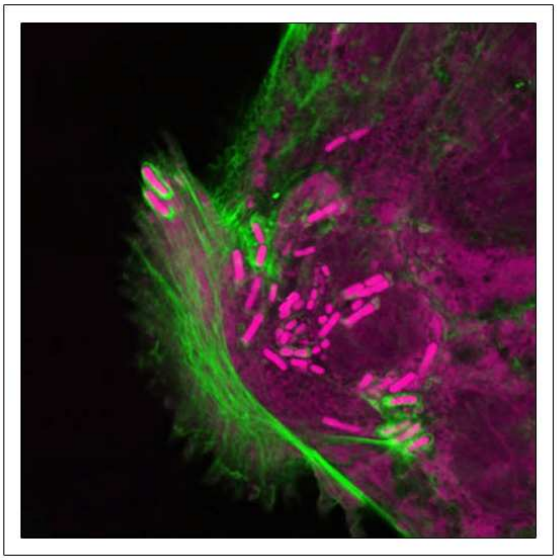
Molecular characterization of an environmental bacterial isolate with strong serological cross-reaction with *Shigella boydii* 15

Implications for live shigellosis vaccine development

by
Eirin Lunden Jenssen



Master thesis submitted in partial fulfillment of the requirements for the degree Master of Science in Biology with specialisation in Microbiology

Department of Biology
The University of Bergen
April 2009

Forord/Acknowlegments

The work of this master thesis was performed at the Department of Biology at the University of Bergen and the University of Dhaka between August 2007 and April 2009.

Jeg vil først og fremst gi en stor takk til min veileder professor Nils-Kåre Birkeland. Mellom alle reisene har jeg vært en av de første til å ”booke” tid hos deg og du har vist stor tålmodighet og gitt mye verdifull veiledning, ikke minst i forbindelse med skrivingen av oppgaven. Til og med da jeg i et forvirret øyeblikk på første veiledning spurte ”Hva er proteomikk?” gav du meg ikke opp. Jeg vil også takke masse for at du fikk arrangert en uforglemmelig opphold i Bangladesh.
I forbindelse med oppholdet i Bangladesh vil jeg også takke MSc Ziaur Rahman, MSc Nafisa Azmuda og professor Sirajul Islam Khan for den gode mottakelsen, veiledning og hjelpen med oppgaven.

Et enormt takk må også gis til vår unike avdelingsingeniør Dr Marit Madsen. Din hjelp gjennom alt labarbeidet har vært helt uunnværlig. Du holder orden og har en stor tålmodighet med oss ferske labarbeidere, og har reddet mangt et katastrofalt eksperiment på laben.

Som ”fersking” ved UiB var det veldig fint komme inn i et godt sosialt miljø på mikrobiologisk både i labgruppen og i gjengen på lesesalen. Det har til tider vært skremmende å se utviklingen vi har gjort fra normale studenter til særinger som ikke klarer å møtes uten å snakke mikrobiologi.

En stor takk vil jeg også gi til mamma og pappa som helt uten krav har støttet meg gjennom hele prosessen både moralsk og økonomisk når lånekassen har vært mindre samarbeidsvillige. En spesiell takk vil jeg også gi til søster og bestevenn Silje. Du er et funn!

Til slutt vil jeg gi en helt spesiell takk til Johan. Jeg tror du har tatt all oppvasken siden oktober da armene mine kapitulerte. Jeg ser gjerne at denne trenden fortsetter ☺. Du har en merkelig evne til å få meg på godt humør og sette ting i perspektiv som har vært til stor støtte under sykemeldinger og utsettelser.

Bergen, April 2009

Eirin Lunden Jenssen

Frontpage:
Shigella bacteria (bright pink) infecting human intestinal cells. Confocal micrograph by Stephanie Schuller (www.wellcome.ac.uk/en/bia/images/26.jpg).

Summary

Bacillary dysentery caused by *Shigella* bacteria (shigellosis) is a major health problem in developing countries. Even though contaminated drinking water is expected to play an important role for the dissemination of this infectious disease the survival and distribution of *Shigella* bacteria in the aquatic environment is not well understood.

This project focused on a bacterial isolate, called Iso10, from a lake outside Dhaka in Bangladesh. It strongly cross- reacts serologically with *Shigella boydii* 15 specific antisera but does not possess the genes associated with *Shigella* virulence, which makes it a potential candidate for development of a live shigellosis vaccine. The main goals were to identify the relationship between Iso10, *Shigella* and *Escherichia coli* using Multi Locus Sequence Analysis and comparison of the proteomic and LPS characteristics of Iso10 with *S. boydii* 15 trough 2D gel analysis and Western blotting.

The results from sequencing of the 16S rRNA gene and 7 housekeeping genes (p*urA, adk, icd, fumC, recA, mdh* and *gyrB)* from Iso10 showed that it shares significant sequence identity with the species *Esherichia fergusonii.* A phylogenetic tree based on concatemers of the 7 genes showed that Iso10 is placed together with *E. fergusonii* outside the main clusters of *E. coli* and *Shigella* spp.

Membrane proteins were successfully extracted from Iso10 and separated on a 2D gel. Ten different proteins were identified by deNovo sequencing (MALDI TOF-TOF MS) and Peptide Mass Fingerprinting with MALDI-TOF MS.

One protein, showing a very strong cross- reaction with *Shigella*- specific antisera, was identified as Outer membrane protein A (OmpA). A comparative study with OmpA extracted from Iso10, *E. coli* K12, *S. boydii* 15 and *E. fergusonii* showed that all the proteins reacted with *Shigella*- specific antisera, indicating that OmpA is a highly conserved porin in the *Escherichia/Shigella* group. OprF, a membrane protein previously only detected in *Pseudomonas* spp., was identified by deNovo sequencing which indicates an exchange of antigens between species which has not earlier been observed.

LPS from Iso10, *E. fergusonii* and *S. boydii* was extracted and probed with *S. boydii* 15 type specific antisera. The results showed that Iso10 and *S. boydii* cross- reacted but no cross-reaction could be detected between *E. fergusonii* and *Shigella*- anitsera, which indicates a previous horizontal gene transfer of the O- antigen gene cluster (*rfb*) between *S. boydii* 15 and Iso10. Iso10 may induce an immune response capable of protecting against *S. boydii* 15 infections. However, since the virulence of *E. fergusonii* still is unclear further characterization has to be carried out before it can be evaluated as a possible live vaccine.

1 Introduction

1.1 Background and objectives for the study

The *Shigella* bacteria is the causative agent of shigellosis or bacillary dysentery, a disease responsible for major health problems in developing countries today ranging from mild diarrhoea to severe dysenteric syndrome (Maurelli and Sansonetti 1988). The annual number of shigellosis episodes is estimated to be 160 millions with 1.1 million deaths, mainly children under 5 years old (Kotloff et al. 1999). Due to a low infection dose (10 cells) *Shigella* is easily spread in areas with poor sanitation and low hygiene by direct fecal–oral route or by contaminated food, flies and water (Hale 1991).
Recently, the emergence of multi– resistant *Shigella* strains is reducing options for effective antibiotic treatment. This is complicating the treatment in developing countries with poor health systems, and increases the need for an effective *Shigella* vaccine (von Seidlein et al. 2006).

Shigella studies are normally based on isolates from the stools of diarrhoeal patients. Even though drinking water is thought to be essential for the dissemination of *Shigella,* only one report has earlier described the isolation of *Shigella* from environmental water (Faruque et al. 2002).
As a part of the NUFU project "Assessment of microbial pollution, diversity and community structure in freshwater resources in Bangladesh using molecular techniques" the Department of Microbiology at the University of Dhaka has succeeded in isolating eleven bacterial strains from ponds, lakes and rivers around Dhaka city by following novel protocols designed and optimized for isolation of *Shigella* (Rahman et al. 2007). The strains form *Shigella*- like colonies on selective agar and show serological cross-reaction with *Shigella*-specific antisera. This master project focuses on one of the environmental isolates, called Iso10, which cross-reacts strongly with antisera from *Shigella boydii* 15. This specific isolate is of special interest because it does not endure any of the virulence genes present in *Shigella boydii* and can therefore be a tool in the development of a live *Shigella* vaccine.

The master project is divided into three parts:

1) Multi Locus Sequence Typing (MLST) analysis of seven housekeeping genes and 16S rRNA gene sequencing of strain Iso10.
2) 2D gel of soluble and membrane protein from Iso10 compared with *Shigella boydii* 15 type strain. Identification of proteins by DeNovo sequencing and Peptide Mass Fingerprinting (PMF).
3) Serological comparisons using Western analysis of protein and Lipopolysaccharides (LPS) gels.
   Western blot of 2D gels/SDS– PAGE gels with membrane proteins and LPS from Iso10 and *Shigella boydii* 15 probed with *S. boydii* 15 type- specific antisera.

1.2 Worldwide disease burden of shigellosis

The *Shigella* genus is divided into four species: *Shigella dysenteriae, Shigella flexneri, Shigella boydii and Shigella sonnei.* Each of the species is subdivided into several different serotypes except *S. sonnei* which only have one serotype. The serogroups are irregularly distributed worldwide but industrial countries seem to have a similar pattern of outbreaks. A study performed by Kotloff et al. (1999) estimated an annual number of 164.7 million shigellosis episodes worldwide, of which 163.2 were in development counties and 1.5 million in industrialized countries.

During the last 50 years pandemic waves if Shiga dysenteri (*S. dysenteiae* type 1) have hit South East Asia, Africa and the Indian subcontinent as a consequence of natural disasters and war, placing populations in vulnerable situations in overcrowded areas with poor sanitation and low water quality (Bennish et al. 1990). *S. dysenteriae* 1 usually leads to the most severe stages of dysentery and life-threatening complications and produce a high level of Shiga toxin, which is only produced at low level in *S. sonnei* and *S. flexneri* (Keusch and Jacewicz 1977). Even though the *S. dysenteriae* 1 infection has the most dramatic outcome of *Shigella* infections it is not the dominating sereogroup compared to the total number of *Shigella* episodes. In developing countries the median percentage of isolates of *S. flexneri, S. sonnei, S. boydii* and *S. dysenteriae* is, 60%, 15%, 6% and 6%, respectively, but in industrialized countries 16%, 77%, 2% and 1%, respectively (Figure 1-1) (Kotloff et al. 1999).

Outbreaks in Western countries seem to occur sporadically and often involve uncooked food. Endemic shigellosis caused by direct fecal–oral contamination in industrialized countries

usually involves *S. sonnei* and is often connected to mental hospitals, day care centres, and prisons (Hale 1991). High risk groups also involve travellers, overseas soldiers and homosexual men (Kotloff et al. 1999).

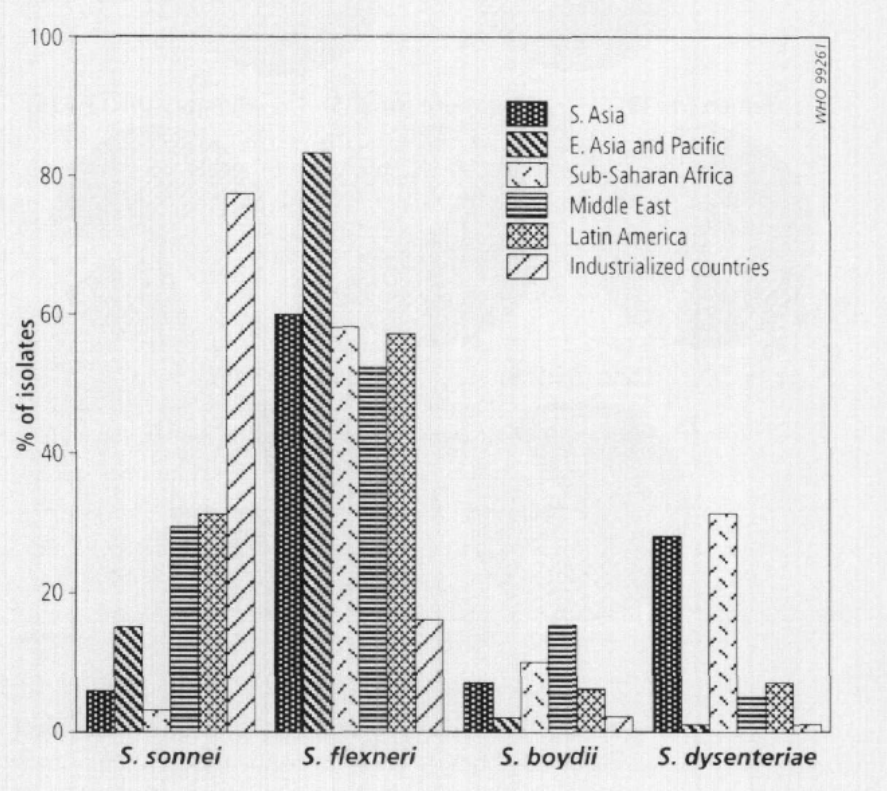


Figure 1-1: Percentage of *Shigella* isolates belonging to four serogroups, by region (Kotloff et al. 1999).

In developing countries *S. flexneri* is clearly the dominating subgroup. This data is supported by a newer study (von Seidlein et al. 2006) of *Shigella* outbreaks in six Asian countries (China, Indonesia, Vietnam, Pakistan, Bangladesh and Thailand). *S. flexneri* was the most frequent serotype isolated (68%) in all the countries except Thailand. Thailand has been undergoing rapid industrialization the last 30 years and here, like in other industrial countries, the most common serotype was *S. sonnei* (85%). *S. dysenteriae* was infrequently isolated in this study (4%) and none of the isolates were identified as *S. dysenteriae* type 1.
Curiously, the rarely detected subgroup *S. boydii* represented nearly one quarter (23%) of the total episodes of *Shigella* in Bangladesh (von Seidlein et al. 2006).
This study also revealed a significant shift in the frequency of the different serotypes of *S. flexneri* both geographically and temporally. This may complicate a future development of a *Shigella flexneri* vaccine since only type-specific immunity has been detected in humans (Mel et al. 1965).

Antimicrobial resistance has prevented efficient treatment of shigellosis since the 1940s when resistance against sulphonamide first was detected in Japan. Ampicillin became the first choice for treatment in the 1950s after the emergence of *Shigella* strains resistant to tetracycline and chloramphenicol. Due to extensive use, ampicillin-resistant strains emerged in Africa and Asia during the 1980s. After years with inappropriate use of antibiotics and over counter sales without prescriptions *Shigella* is developing further resistance against antibiotics (Niyogi 2007)(Table 1-1). In 2004 The World Health Organization recommended ciprofloxacin, which earlier had been used as a back–up drug for cotrimoxazole and nalidixic acid, as the first of choice drug (WHO 2005). Studies of shigellosis episodes in China revealing 100% resistance among *S. flexneri* and 98% resistance among *S. sonnei* isolates against nalidixic acid shows that alternative drugs already is needed in special areas of developing countries (von Seidlein et al. 2006).

Table 1-1: Antimicrobial resistance pattern for recent clinical *Shigella* spp. isolates from 6 different Asian countries (China, Indonesia, Vietnam, Pakistan, Bangladesh and Thailand) (adapted from von Seideln, Kim et al 2006).

Antibiotic	***S. flexneri* (% of n) (n=1976)**	***S. boydii* (% of n) (n=189)**	***S. dysenteriae* (% of n) (n=110)**	***S. sonnei* (%of n) (n=652)**
Ampicillin	1653(83.7%)	46 (24,3%)	41 (37,3%)	64 (9,8%)
Cotrimoxazole	1490(75,4%)	91(48,2%)	68 (61,8%)	603 (92,5%)
Nalidixic acid	473 (23,9%)	40 (21,2%)	9 (8,2%)	92(14,1%)
Ciprofloxacin	31(1,6%)	2 (1,1%)	0 (0%)	1 (0,2%)
Multidrug resistant*	18 (0,9%)	0 (0%)	0 (0%)	0 (0%)

* Isolates are resistant to ampicillin, cotrmoxazole, nalidixic acid and ciprofloxacin.

1.3 Current status of Shigella vaccine development

The development of a safe and effective *Shigella* vaccine would be a major contribution to the public health system to fight the spread of shigellosis. Several vaccine candidates are under development, but currently there is no licensed vaccine against *Shigella* outside China (Venkatesan et al. 1991). *Shigella* has 47 different serotypes. Since only serotype-specific immunity has been observed in humans a polyvalent vaccine representing the most important subspecies and serotypes of *Shigella* is needed. A polyvalent vaccine which includes 100% of *S. flexneri* serotypes, *S. dysenteriae type* 1 and *S. sonnei* (consist of only one serotype) would cover 79% of *Shigella* infections in developing countries and 83% in industrialized countries (Kotloff et al. 1999).

Candidate vaccines currently under development include polysaccharide conjugate and live attenuated vaccines (WHO Updated February 2009). The polysaccharide conjugate vaccine is based on lipopolysaccharides (LPS) from the relevant serotypes and conjugation to tetanus toxid or recombinant *Pseudomonas aerguinosa* exotoxin A (rEPA). A phase III efficacy trial of a *S. flexneri* 2a and *S. sonnei* conjugate polysaccharide vaccine was recently carried out in Israel (Passwell et al. 2001). An example of a live attenuated vaccine is the bivalent vaccine currently only licensed in China based on a Shigella strain (T_{32}-Istrati) expessing both *S. flexneri* 2a and *S. sonnei* O-antigens with deletions in the invasion plasmid and virG/icsA genes. However, a major drawback with this potential vaccine is the three dose regimen with high doses of live vaccine ($2x10^{10}$cfu) (Kweon 2008).

1.4 The virulence of the Shigella in historical perspective

Initially, *Shigella* was described as *Bacillus* (now *Escherichia*) *dysenteriae* by Shiga in 1898. After thorough characterisation of the genus, the 1950 "Congress of the International Association of Microbiologists *Shigella* Commission" recommended that *Shigella* was to be adopted as the generic name with four subspecies: *S. dysenteriae*, *S. boydii*, *S. flexneri* and *S. sonnei* (Hale 1991).

Further research during the 1950s and 1960s employed the close genetic relationship between *Shigella* and *E. coli* and was concentrated on the genes involved in *Shigella* virulence. Experiments using conjugation between *E. coli* K─12Hfr donor and *S. flexneri* recipient revealed several loci that are important for *Shigella* virulence, among these the *xyl–rha* region coding for the ability to multiply intracellularly (Maurelli and Sansonetti 1988).

In 1964 it was discovered that one of the essential steps in pathogenesis of shigellosis is the invasion of the human colonic mucosa (Labrec et al. 1964).

Later, as a part of a program to develop live oral vaccine against shigellosis, experiments using a *S. flexneri* Hfr donor and an *E. coli* K─12 recipient to confer *Shigella* virulence, showed that even when all essential virulence genes were transferred, *E. coli* failed in conferring a *Shigella* pathogenicity phenotype (Formal and Hornick 1978).

A breakthrough came in the early 1980s when Sansonetti and his science group revealed that a plasmid present in *S. sonnei* invasive form I strains was absent in non-invasive form II strains. When this plasmid was introduced to the non-invasive form II strain, the bacteria expressed the invasive phenotype and somatic specificity of the donor strain (Sansonetti et al. 1981).

A similar plasmid was later detected in *S. flexneri* and then extended to include the other species of *Shigella* and enteroinvasive *E. coli* (Sansonetti et al. 1983).
Genetic studies have since then revealed that several plasmid genes are necessary to express the pathogenic phenotype. Multiple regulatory loci in the chromosome and plasmid that play an important part in the *Shigella* virulence phenotype have been identified and sequenced (Hale 1991).

1.5 Molecular relationship of E. coli and Shigella species

Even though *E. coli* and *Shigella* spp. are classified as distinct species, they have always been considered to be very closely related. In the 1940-50s *E. coli* isolates were described which tested positive for the Sereny test by causing experimental keratoconjunctivitis in guinea pigs. The invasion and multiplication in corneal epithelium mimics *Shigella* behaviour in intestinal epithelium. A positive Sereny test is often used as a characteristic to identify *Shigella* (Maurelli and Sansonetti 1988). The strains were first classified as *Shigella*, but were later put in a new *E. coli* subgroup called enteroinvasive *E. coli* (EIEC) (Lan et al. 2004).
There are few biochemical and physiological characteristics that separate EIEC from *Shigella*. Most *Shigella* and EIEC strains are identified by O- serotyping with commercially available antisera, but even some of the O antigens associated with EIEC are identical to those found in *Shigella* spp. (Coimbra et al. 2000).

Since the mid–eighties there has been an ongoing discussion concerning two aspects of the *Shigella –E. coli* relationship. Should *Shigella* be defined as a group within *E. coli* and has *E. coli* arisen many times or evolved from one common *E. coli* ancestor.
The later years it has been more or less accepted that *Shigella* does not form a discrete group within *E. coli* and do not even represent a subspecies. Already in the 1980s studies using Multilocus Enzyme Electrophoresis (MLEE) showed that *Shigella* and *E.* coli could not be separated into two different genera (Ochman et al. 1983). A later study which included all the different pathogenic groups of *E. coli* and *Shigella* concluded that *Shigella* fall into different clusters of *E. coli* and does not form one discrete group (Pupo et al. 1997). A newer project including all the serotypes of *Shigella* was performed by sequencing eight housekeeping genes. It identified three main clusters of stains all including the strains from more than one of the traditional subgroups and five forms not closely related to any of the other (Pupo et al.

2000). On the basis of these phylogenetic analyses it was strongly argued that *Shigella/E. coli* have derived from multiple independent origins. This was supported by the major phylogenetic study of *Shigella/E. coli* by Wirth et al (2006). However, a different study of the virulence plasmids of *Shigella* and EIEC strains showed that the chromosome and the virulence plasmid share a similar evolutionary history. From this the researches concluded that the virulence plasmid was transferred into an ancestral *E. coli* that gave rise to all the *Shigella* and EIEC groups (Escobar-Paramo et al. 2003).
This is still an ongoing discussion and none of the theories are fully accepted. However, recently it seems like more and more groups of scientists support the multiorgin theory (Yang et al. 2007).

1.6 Phylogenetic studies of Shigella and E. coli

In 1984 72 *E. coli* isolates from different geographic locations and from a variety of human and animal hosts were chosen to represent the known genetic diversity of *E. coli* (Ochman and Selander 1984). This is called the *E. coli* reference collection (ECOR). MLEE studies of the ECOR collection divided *E. coli* into 4 subgroups: A, B1, B2, D and a minor group E (Herzer et al. 1990) (Figure 1-2).
In 2006 a new database based on sequences of seven housekeeping genes from 460 isolates of *E. coli* (including *Shigella*) from a number of different sources was established (Wirth et al. 2006). This database, established by Mark Achtman and now known as the MLST database for *E. coli* at Environmental Research Institute (ERI) (University College Cork), was used in a study which revealed a new subdivision of *E. coli* strains by using population genetic tools. In addition to subgroup A, B1, B2 and D they defined one new hybrid group named AxB1, and a second group called ABD. The new groups included isolates which had gone through extensive recombination between the different classical ECOR groups (Figure 1-3).
A tree defining the six ancestry groups was presented and compared with five of the pathogenic groups of *E. coli* (Figure 1-4 and 1-5). This showed that over 50% of the EIEC and *Shigella* strains fell into hybridgroup ABD and AxB1, indicating a connection between virulence and homologous recombination.

Figure 1-2: Heuristic maximum likelihood tree based on neighbour joining (NJ) including concatenated sequences of the 7 housekeeping genes from the ECOR collection. The numbers at the nodes are bootstrap values above 70% (Wirth et al. 2006).

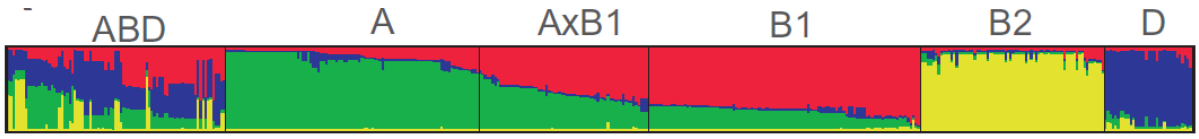


Figure 1-3: Propotions of ancestry from groups A, B, B2, and D inferred by STRUCTURE and their assignment to six groups as displayed with DISTRUCT. The plot shows one vertical line for each isolate indicating the proportions of ancestry from the four groups, colour indicated as in Figure 1-4 (Wirth et al 2006).

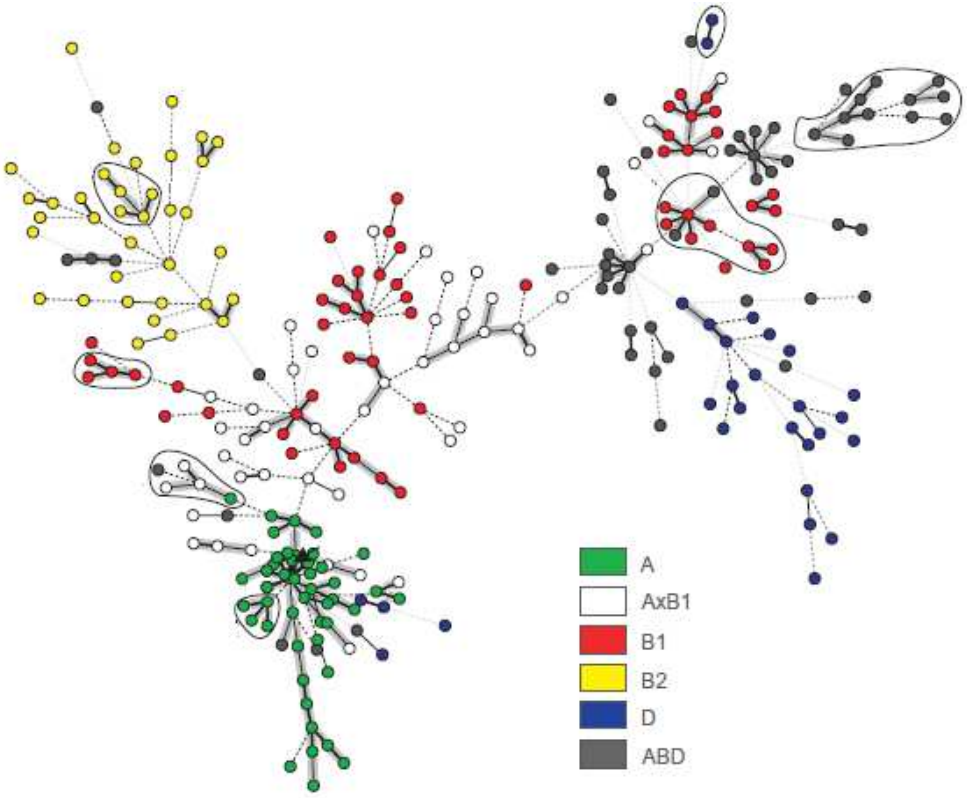


Figure 1-4: Distribution of ancestry from groups A, B1, B2, D, AxB1 and ABD in a minimal spanning tree of 275 *E. coli* Sequence Types (ST)based on degree of allele sharing. Dots, each representing one ST, are coloured according to groups and lines connecting ST complexes are shaded in gray (Adapted from Wirth et al 2006).

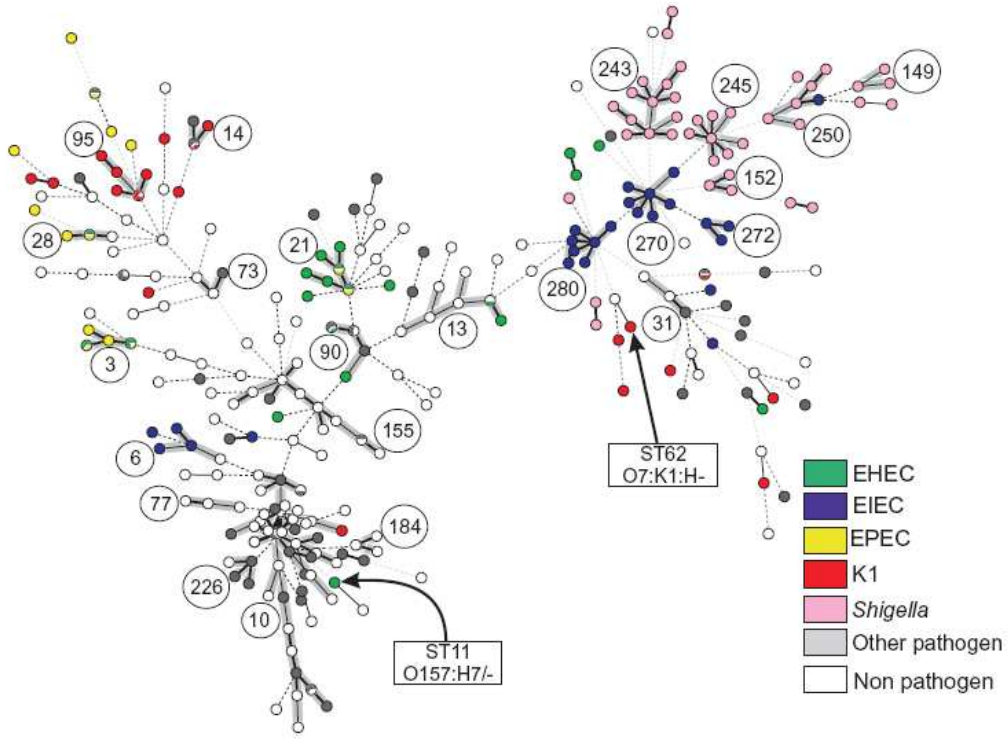


Figure 1-5: Pathogenic types of E.coli presented in a minimal spanning tree. Different colours represent different pathogenic types of E.coli. Circled numbers indicate ST complexes whereas arrows indicate STs 11 and 62. The lines connecting the pairs of ST specify how many allelic pairs they share: Black, thick lines (six alleles shared), black, thin lines (five), black, dotted lines (four), grey, dotted lines (three to one). Gray, shaded lines indicate that the line connect STs with an ST complex (Wirth et al 2006).

1.7 Earlier work with the environmental isolate Iso10

Iso10 has earlier been a central part of the scientific work in one Master thesis by Nafisa Azmuda (Azmunda 2006) and the ongoing doctoral thesis work of Mohammed Ziaur Rahman (Rahman et al. 2007) at the Department of Microbiology, University of Dhaka in Bangladesh. The most important findings in this work will be presented in the following sections.

1.7.1 Morphological and cultural characteristics and antibiotic susceptibility pattern

The sampling and culturing methods used to isolate Iso10 is described in Rahman et al. 2007. By microscopic techniques Iso10 was identified as a rod shaped, non-motile, gram negative bacterium. It grows on selective and differential agar like MacConkey agar, Xylose Lysine Desoxycholate agar (XLD), Hektoen Enteric agar (HEA) and *Salmonella-Shigella* agar (SS). Iso10 was found to be multiple resistant to the following antibiotics: Streptomycin, erythromycin, trimethoprim– sulfamethoxazole and nalidixic acid.

1.7.2 Slide agglutination test for serological identification and Enzyme Linked Immunosorbent Essay (ELISA)

The slide agglutination test is used routinely for serological identification of bacteria. Colonies from a pure culture is mixed with specific antisera on a class slide and observed for agglutination.

Iso10 showed strong agglutination for group specific *Shigella boydii* polyvalent antisera and *Shigella boydii* 15 type specific antisera, but no reaction was obtained with the other type specific antisera.

ELISA is a rapid immunochemical test for detection of antigen-antibody reaction. An ELISA-plate is coated with whole cell extracts from the bacteria. Primary antibody and enzyme linked secondary antibody reacts with the antigen. The final substrate is added and converted by the enzyme to a coloured product which can be measured by a spectrophotometer.

Iso10 gave a very strong reaction in the ELISA test ($1{:}5^7$) with *S. boydii* 15 type specific commercial antisera and yielded even higher titers that the positive control (*S. boydii* type 15; ATCC12034).

1.7.3 Detection of virulence genes and random amplified polymorphic DNA assay (RAPD)

Genomic DNA from Iso10 was subjected to PCR analysis with primers specific for amplification of *Shigella* virulence genes. No PCR product was obtained for genes coding for

ipaH, ipaBCD, virA, ial and *stx*, indicating that Iso10 does not possess these *Shigella* specific virulence genes. The absence of these genes was also verified by Southern analysis.
The principle for RAPD is a PCR reaction where unspecific primers and chromosomal DNA are used to produce a RAPD profile. This will detect small differences in the chromosomal DNA for different isolates. The RAPD profile from Iso10 and *S. boydii* 15 did not match (Figure 1-6).

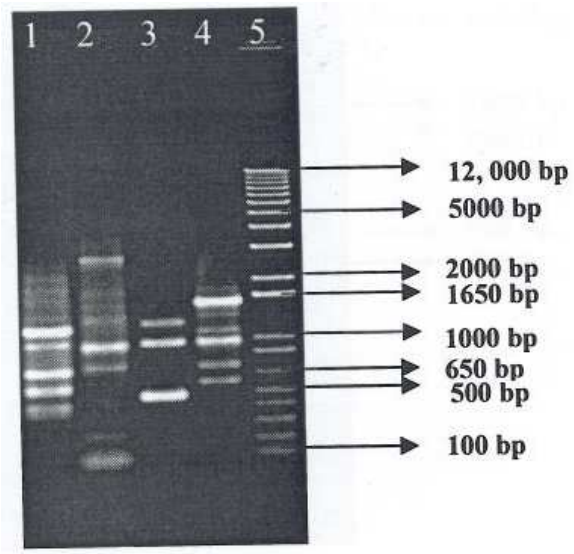


Figure 1-6: Representative RAPD pattern of Iso10 and *S. boydii 15*. Lane 1 and 2 contain the fragments of Iso10 and *S. boydii* respectively analysed with primer set 1. Lane 3 and 4 contains the fragments of Iso10 and *S. boydii* respectively analysed with primer set 2. Lane 5 contains 1kb plus DNA marker.

1.7.4 Restriction fragment length polymorphism analysis (RFLP), ribotyping and restriction analysis of the O antigen gene cluster

Restriction fragment length polymorphism (RFLP) uses restriction enzymes to cut chromosomal DNA and produce a unique RFLP pattern. This pattern can be transferred from an agarose gel to cellulose membrane and hybridized with specific DIG– labelled DNA probes such as a 16SrRNA gene probe (ribotyping). Ribotyping of Iso10 and *S. boydii* (ATCC 12034) showed that only 5 out of 7-9 bands matched (Figure 1-7), indicating a difference in genome organization.

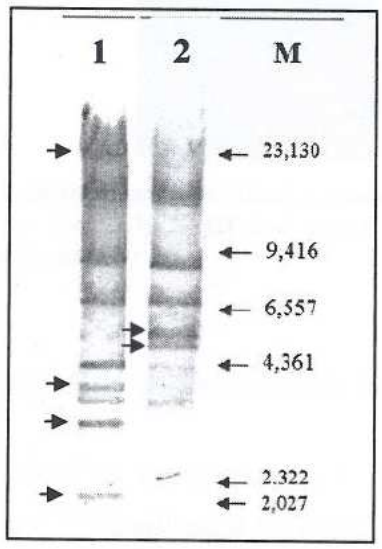


Figure 1-7: Pattern of bands after ribotyping analysis. Lane 1, 2 and M represent the bands of *S. boydii*, Iso10 and λ-DNA /*Hind*III markers respectively. The small arrows indicate bands which are different.

The Lipopolysaccharides (LPS) is a vital part of the outer membrane of Gram-negative bacteria. LPS usually consists of three regions: Lipid A, core oligosaccharide and variable polysaccharide side chains (O antigen). The O antigen is the most variable part, made out of many repeats of O units(Guo et al. 2005). The O antigen gene locus, *rfb,* encoding the O-antigen biosynthesis genes, is located between the *galF* and *gnd* genes, and varies in size and nucleotide sequence, making it possible to separate bacteria into O antigen gene cluster restriction fragment types based on RFLP analysis of this locus (Schnaitman and Klena 1993). The O antigen gene cluster of Iso10 was PCR amplified, and the product was treated with Mbo11 restriction enzyme and analysed by agarose gel electrophoresis. The restriction fragment from Iso10 and *S. boydii* showed a similar pattern on the agarose gel and varied only in three weak fragments (Figure 1–8). This corroborates the strong serological O-antigen cross-reactivity and implies that Iso10 and *S. boydii* type 15 share a significant O-antigen gene homology.

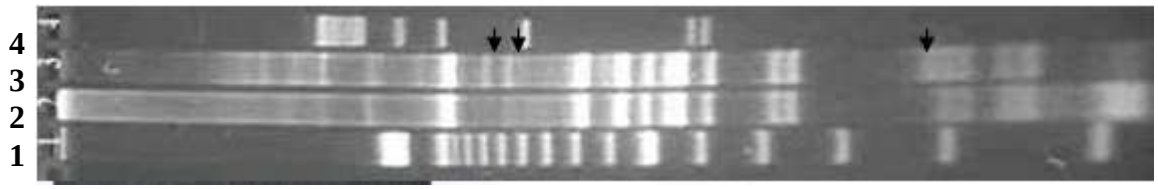


Figure 1-8: The restriction pattern of the O–antigen gene cluster from Iso10 and *S. boydii* in lane 2 and 3 respectively. The small arrows indicate the bands which are different.

1.8 Escherichia fergusonii

During the experimental work with Iso10 it early became clear that the strain showed many similarities with the species *Escherichia fergusonii.* This is a rare member of *Escherichia* that was first known as Enteric group 19, but was proposed as a new species by Farmer in the mid-eighties (Farmer et al. 1985).

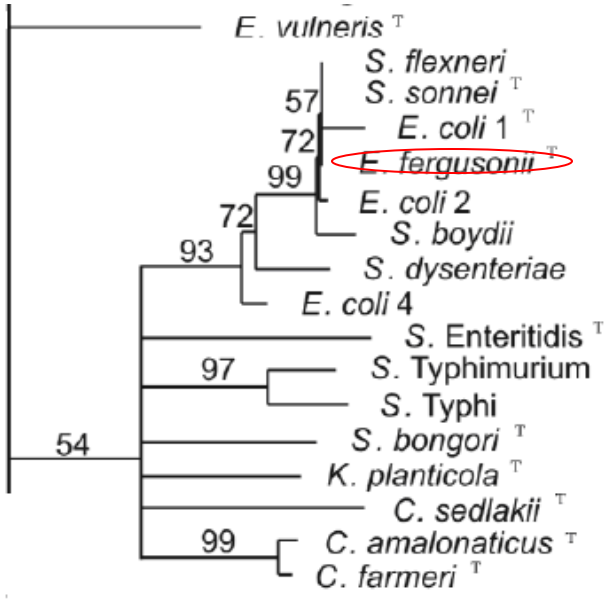


Figure 1-9: Part of a phylogenetic tree of the members of *Enterobacteriaceae* based on sequences from 16S rDNA gene generated by the neighbour joining method. Adapted from Paradism Boissiont et al (2005)

Based on 16S rRNA sequences *E. fergusonii* falls into *E. coli- Shigella* cluster as shown in the phylogenetic three in Figure 1–9 (Paradis et al. 2005).

However, biochemical characterisation and DNA-DNA hybridisation has shown that *E. fergusonii* should be classified as a separate species within genus *Escherichia.*

Biochemically, it can be differentiated from *E. coli* by being sorbitol and lactose negative but adonitol positive. *E. coli-Shigella* is the closest relatives to *E. fergusonii* by DNA–DNA hybridisation, sharing 49-63% similarity (Farmer et al. 1985). This taxa has been confirmed by phylogenetic analyses based on the sequences of outer membrane protein 3A (*ompA)* and glyceraldehyde–3– phosphate dehydrogenase (*gap)* as shown in Figure 1–10 (Lawrence et al. 1991).

E. fergusonii has been isolated from a variety of human and animal clinical specimen including blood, urine and faeces, but the pathogenic potential of *E. fergusonii* is not clear. It has frequently been associated with wound infections but have also been reported from other types of infections (Mahapatra and Mahapatra 2005).

Funke et al. (1993) reported isolation of *E. fergusonii* from four different sites (gall bladder fluid, blood cultures, faeces and wounds) from a patient with pancreatic carcinoma and cholangiosepsis (Funke et al. 1993). It has also been found to cause disease clinically similar to salmonellosis in sheep and cattle by leading to abortion and mastitis (Bain and Green 1999).

In a recent study using an *E. coli* virulence gene array, 30 isolates of *E. fergusonii* were tested for the presence of *E. coli* virulence genes *iss* (increased serum survival), *prfB* (P-related fimbriae regulatory) and *ireA* (sideropore receptor ireA). 43% of the samples tested positive for the *iss* gene, while 10% were *prfB* and *ireA* positive (Wragg et al. 2008). This links *E. fergusonii* isolates to pathogenic types of *E. coli* like ExPEC and APEC.

During routine screening of beef in 2006, a strain of *E. fergusonii* which tested positive for *E. coli* O157 antibodies was observed in Australia. The isolate tested positive for the O157gene by PCR, and further tests revealed a high level of sequence identity and identical gene order between *E. coli* O157 O–antigen gene cluster and the *E. fergusonii* strain. This indicates that the O157 O–antigen has been transferred trough horizontal gene transfer between *E. coli* O157 and the *E. fergusonii* strain (Fegan et al. 2006).

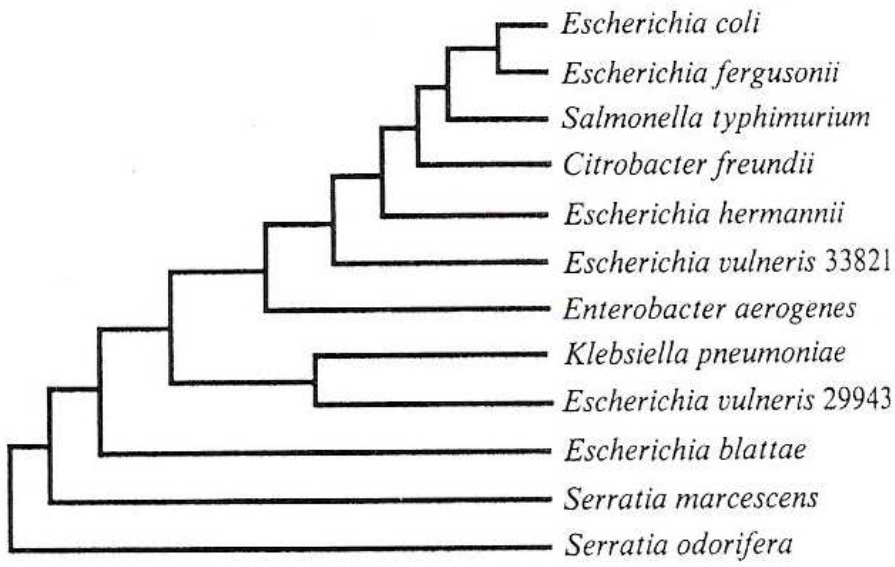


Figure 1-10: Phylogeny of selected enteric bacteria based on the sequences of *gap* and *ompA* genes, generated by a maximum parsimony algorithm (Lawrence et al 1990).

1.9 Multi Locus Sequence Typing

Multi Locus Sequence Typing (MLST) is based on the same principles as its predecessor, Multi Locus Enzyme Electrophoresis (MLEE), but uses the DNA sequences from multiple housekeeping genes to detect differences among bacterial strains. In most cases sequences from 4-8 genes are sufficient to discriminate between bacterial strains (Cooper and Feil 2004). All allelic variants for a given locus are prescribed an allelic number. The combination of the allelic numbers for each locus for a given strain composes the Sequence Type (ST). Many MLST datasets arrange related ST into groups, called ST complexes. There is no general definition of a ST complex but it is normally based on the principle that some STs are more frequent in the dataset than others. An ST complex can for example consist of a group of minimal three sequence types that share at least four out of seven allelic numbers (Urwin and Maiden 2003).

An important part of an MLST analysis is the choice of genes. Cooper and Feil (2004) distinguished between three classes of genes: Informational genes involved in essential functions of the cell (e.g. transcription), core housekeeping genes encoding proteins central in cell metabolism (e.g. enzymes in the TCA cycle) and hypervariable genes typically expressed on the cell surfaces or associated with virulence in pathogenic bacteria.

The 16S rRNA gene is a typical example of an informational gene which is widely used for identification of bacteria. Informational genes evolve extremely slowly, and taxonomic studies based on sequencing of 16S rRNA is in some cases only effective for studies of different genera and to a certain degree species but not intra-species comparisons, like for example in the *Escherichia – Shigella* clade, where the variation in 16S rRNA sequence is too low to distinguish between *Escherichia* and the *Shigella* spp. This was also shown in the phylogenetic tree of the members of *Enterobacteriaceae* (Figure 1-9) where *E. fergusonii* groups with *E. coli* strains.

Housekeeping genes typically evolves at moderate rate, and is preferred in most studies over hypervariable genes in MLST analyses. Taxonomic studies based on genes which evolve at extremely high rate might lead to discrimination of strains which in reality is very closely related (Cooper and Feil 2004). However, in studies where recognition of microevolution is essential, for example identification and separation of strains in a specific geographic area, analysis of hyper-variable genes can be very useful (Maiden et al. 1998).

The MLST technique was first used in 1998 by Maiden and his staff with strains of *Neisseria meningitides* (Maiden et al. 1998). During the last ten years numerous online MLST databases including a wide range of species have been developed. In 2004 the MLST database for *E. coli* (including *Shigella*) at Environmental Research Institute (ERI), University College Cork, established by Mark Achtman, became available online (http://mlst.ucc.ie/mlst/dbs/Ecoli). This database is based on seven housekeeping genes evenly distributed around the *E. coli* chromosome: *purA, adk, icd, fumC, recA, mdh* and *gyrB* (Figure 1-11) (Wirth et al. 2006). After updating the database in April 2007 it included 600 STs comprising 54 ST complexes. The definition of an ST complex was also changed to only include groups of at least three STs sharing six alleles in pair-wise comparisons.

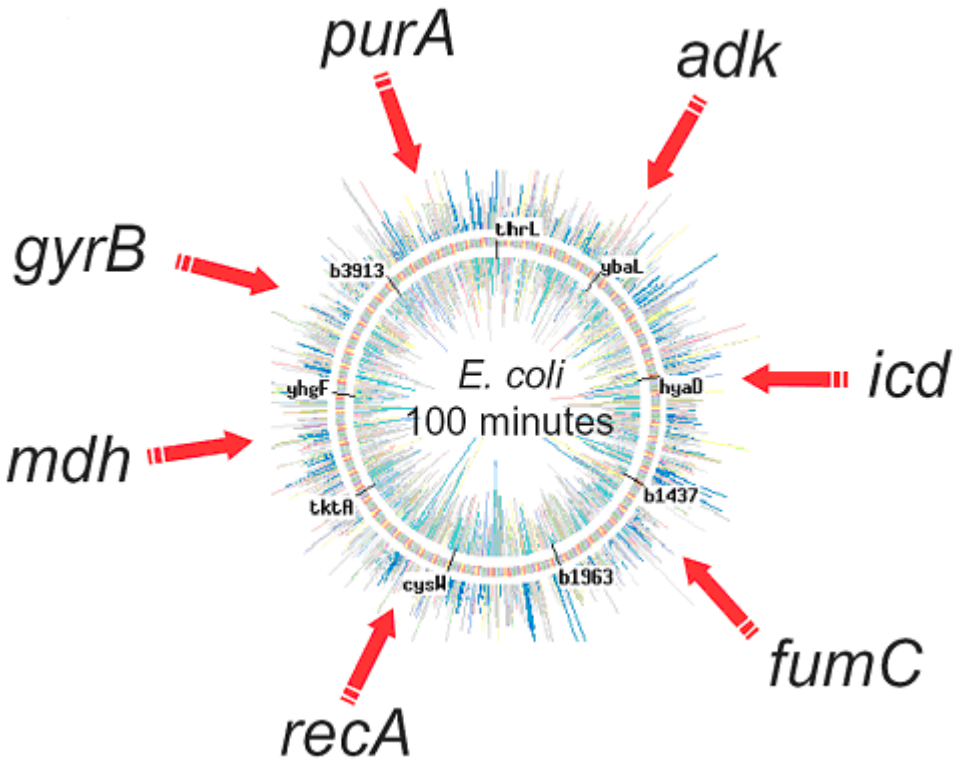


Figure 1-11: The genomic locations of the housekeeping genes included in the Achtman MLST database for *E. coli* (Wirth et al. 2006).

1.10 Proteomic analysis of the bacterial outer membrane

The envelope in gram negative cells makes up a complex structure consisting of the outer membrane, periplasma-peptidoglycan layer and the inner membrane (Figure 1-12). The outer membrane is made up of phospolipids, membrane proteins, lipoproteins and lipopolysaccharides (LPS). Lipoproteins (Lpp or Braun's lipoprotein) is one of the most

abundant membrane proteins connecting the peptidoglucan layer to the outer membrane by its hydrophobic end while other membrane proteins function as porins or hydrophilic channels which allows free diffusion of small molecules or specific transport across the outer membrane (Weiner and Li 2008).

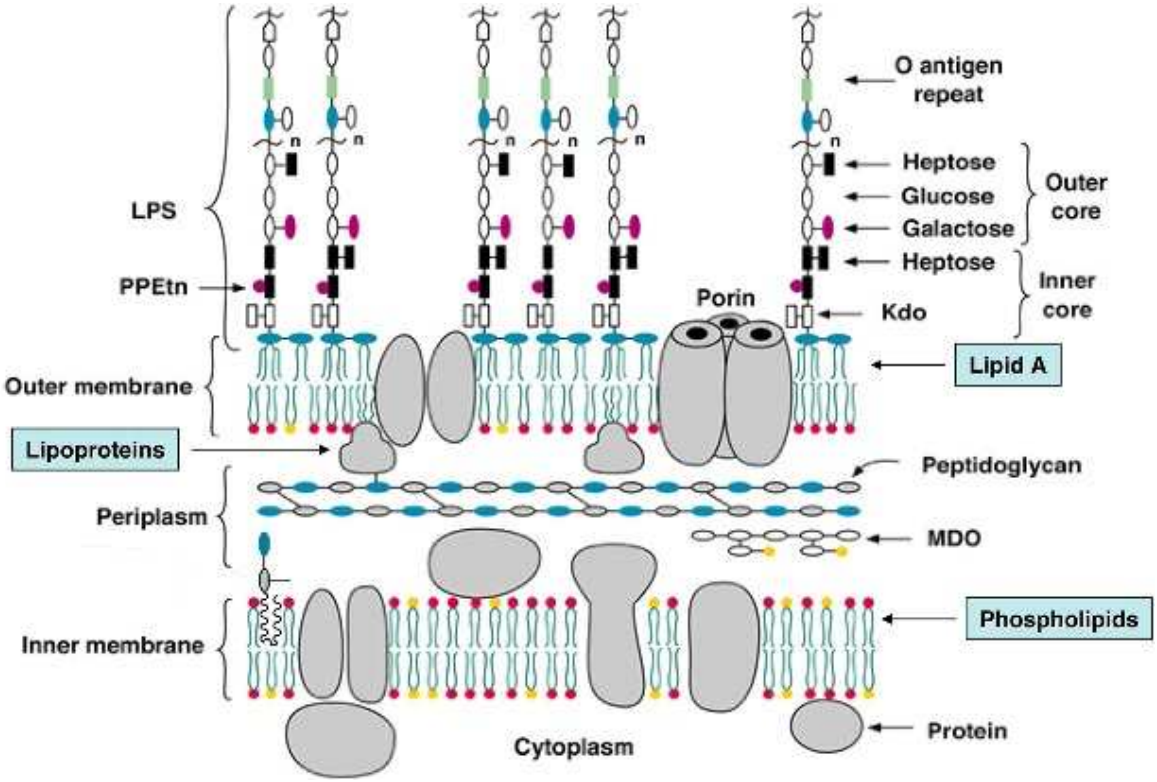


Figure 1-12: Model of the envelope of *E.coli* K12 showing the membrane proteins, lipoproteins and LPS which make up the outer membrane (Raetz and Whitfield 2002)

Two- dimensional gel electrophoresis (2-DE) has during the later years become one of the most important techniques in proteomics. It was developed by O'Farrell and his team in the 1974 to separate soluble proteins of *E. coli.* The method made it possible to separate proteins according to isoelectric point (pI) and molecular weight (M_r) and identify protein spots of interest by Mass Spectrometry (MS) (O'Farrell 1975).

The working process of 2-DE identification can be separated into five steps

1. Protein isolation and solubilisation
2. Protein separation with isoelectric focusing and SDS- PAGE
3. Protein detection and quantification
4. Computer analysis of 2-DE pattern.
5. Protein identification and characterization

1.10.1 Protein isolation and solubilisation

This procedure fractionate soluble and membrane proteins. A widely used technique to isolate soluble proteins is TCA/aceton precipitation. This is an effective method to remove interfering compounds (e.g. salt and polysaccharides) and reduce protein degradation (Gorg et al. 2004).

In the late 90s a study by Wilkins et al. (Wilkins et al. 1998) examined the identified proteins from three species (*E. coli, Bacillus subtilus* and *Saccharomyces cerevisiae)* on 2D gels and compared them to proteins predicted from genome sequences. It revealed that hydrophobic proteins were more or less absent from 2D gels based on whole cell lysates. They also found that the mode of IEF (CA or IPG) did not affect the results but suggested that the extraction and solubilisation of the proteins were the crucial steps to identify hydrophobic proteins.

An effective method for isolation of membrane proteins was developed by Fujiki et al. in 1982 which observed that closed vesicles was converted to open sheets of membranes when they were treated with 0.1M sodium carbonate. As a consequence, peripheral membrane proteins were released in soluble form (Fujiki et al. 1982). This method was modified by Molloy et al. to be adjusted for the 2-DE technique (Molloy et al. 2000).

Solubilization and complete denaturation of proteins is a critical step on the way to achieve a well focused 2D– gel. It is essential to obtain complete disruption of all molecular interaction to guarantee that each spot represents a single polypeptide.

When classic SDS–PAGE is used without isoelectric focusing heat-treatment with the anionic detergent SDS is sufficient to unfold the proteins and break noncovalent interactions. Isoelectric focusing requires that the original surface charge is maintained. Charge- shifting reagents like SDS should therefore not be used (Molloy 2000).

Different components are required depending on the chemical character of the proteins including chaotropes, reducing agents and surfactants (detergents) (Molloy 2000).

Chaotropes*:* Chaotropic reagents break hydrogen bonds, which cause unfolding and denaturation of proteins. The most common chaotrope is urea. Urea is uncharged and normally used in the highest possible concentration (8M) in 2D electrophoresis. Another chaotrope, thiourea, is more effective in breaking hydrophobic interactions and is commonly used together with urea (7Murea + 2M Thiouea). This combination has proved to give higher solubilization, especially for membrane proteins (Molloy 2000; Gorg et al. 2004).

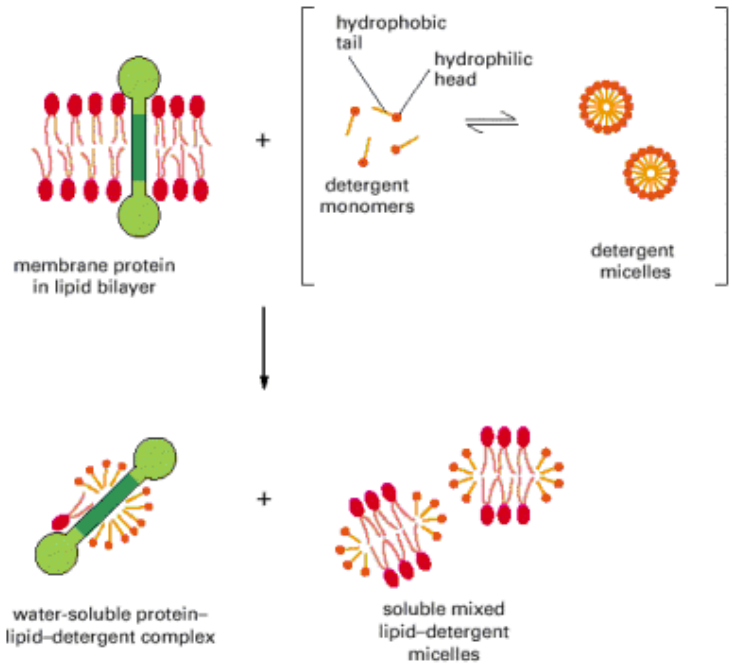


Figure 1-13: Solubilization of membrane proteins in a lipid layer with detergent (Bruce Alberts 2007).

Detergent: Detergents are added to a solution with chaotropes to act together to improve sample solubilization. They also prevent aggregation caused by hydrophobic interaction. This is especially important in solubilization of membrane proteins where detergents break into the lipid bilayer, bind to hydrophobic residues, and bring the proteins in solution by making protein-detergent complexes (Figure 1-13) (Molloy 2000; Bruce Alberts 2007). Both nonionic (e.g. Triton X-100) and zwitterionic (e.g. CHAPS, ASB-14) detergents can be included in the soulibilization mix. Because not all nonionic detergents are effective on very hydrophobic proteins, zwitterionic detergents are often preferred to solubilize membrane proteins(Molloy 2000).

Reducing agents*:* Reductants are included in the solubilization solution to break intra- and intermolecular disulphide bridges. The most common reducing agents are dithiothreitol (DTT) and dithioerythritol (TBP) (Gorg et al. 2004).

1.10.2 Isoelectric focusing

Isoelectric focusing (IEF) is an electrophoretic technique to separate proteins according to their differences in electric charge. Amphoteric substances such as proteins carry positively (NH_3^+) and negatively (COO^-) charged groups. The sum of these groups is called the net charge, and can be positive, negative or zero depending on the surrounding pH (Figure 1-14)

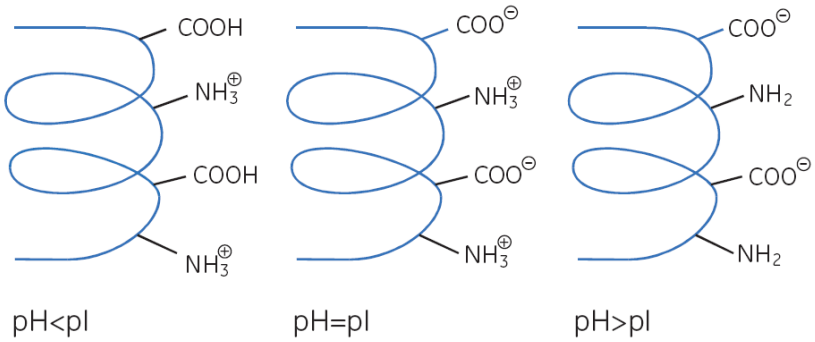


Figure 1-14: The isoelectric point (pI) is the specific pH at which the net- charge of the protein is zero (GEHealthcare 2004).

The isoelectric point (pI) is the specific pH– value at which the protein has an overall net charge of zero. When the net charge of the protein is positive the pH– value is below its isoelectric point. If the pH value increases beyond the isoelectric point of the protein the net charge turns negative (Figure 1-15).

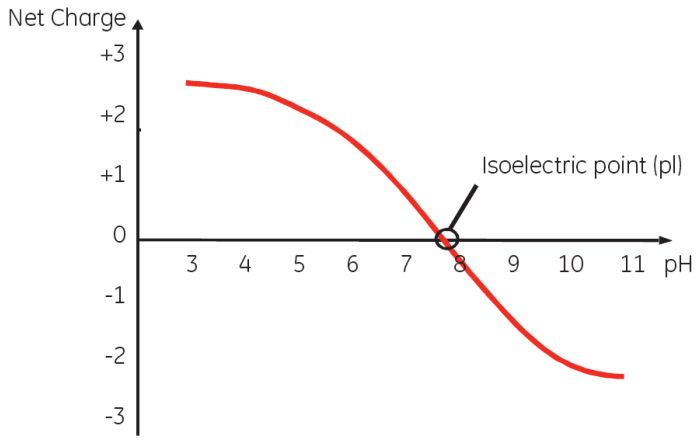


Figure 1-15: Plot of the net charge of a protein versus the pH of its environment. The point of intersection of the curve at the x-axis represents the isoelectric point of the protein (GEHealthcare 2004).

These properties of the proteins are used for isoelectric focusing in a gel with pH gradient under the influence of an electric field. A negatively charged protein will migrate towards the anode until it reaches its isoelectric point and have no electric net charge. Correspondingly, a positively charged protein will migrate towards the cathode and gradually becoming less charged until it reaches its pI. If a protein should diffuse away from its pI it will immediately become charged and migrate back.

Originally, IEF was performed on carrier– ampholyte-generated pH gradients. These are small amphoteric molecules with high buffering capacity at their pI. Under an electric field a

mixture of different carrier ampholytes will arrange themselves according to their pI and form a continuous pH– gradient. The main weakness with this technique was the pH gradient, which tended to be unstable over time leading to cathodic drifting and batch to batch variability. By introducing immobilized pH gradients (IPG) with chemically immobilized buffering and titrant groups, which is not able to migrate in an electric field, many of these issues were overcome (Gorg et al. 2004). IPG –strips are produced by incorporating Immobilines in the polyacylamide gel. Immobilines are not zwitterionic but are small substances with acidic or basic buffering groups. Immboilines IPG–strips were used in this study and followed by SDS-PAGE for separation of proteins according to molecular weight.

1.10.3 Protein detection and computer analysis

2-DE has the later years become an essential technique in comparative studies of the protein expression of prokaryotes. The gels are traditionally stained by anionic dyes (e.g. Coomassie Blue) or Silver staining. The method allows both qualitative comparison by observation of new or lost spots and quantitative comparison by increased or decreased spot intensity.
By using a digital image analysis system, the gel of interest can be aligned, corresponding spots can be paired and differences in protein expression identified.

1.10.4 Protein identification

Mass Spectrometry is a central technique for all proteomic identification. This step was performed by the staff at the PROBE platform (The Proteomic Unit at the University of Bergen) but a basic introduction to the technique is described in the following chapter.
By treating a protein with a specific protease (usually trypsin) that cuts C-terminal to all *Lys* and *Arg* residues, a distribution of peptide fragments with specific peptide masses called the Peptide Mass Fingerprint (PMF) is obtained. The molecular weight of these fragments can be identified by comparison with theoretical peptide mass values in a database.
MALDI –TOF (Matrix Assisted Laser Desorption Ionisation – Time of Flight) is a commonly used Mass Spectrometer for PMF.

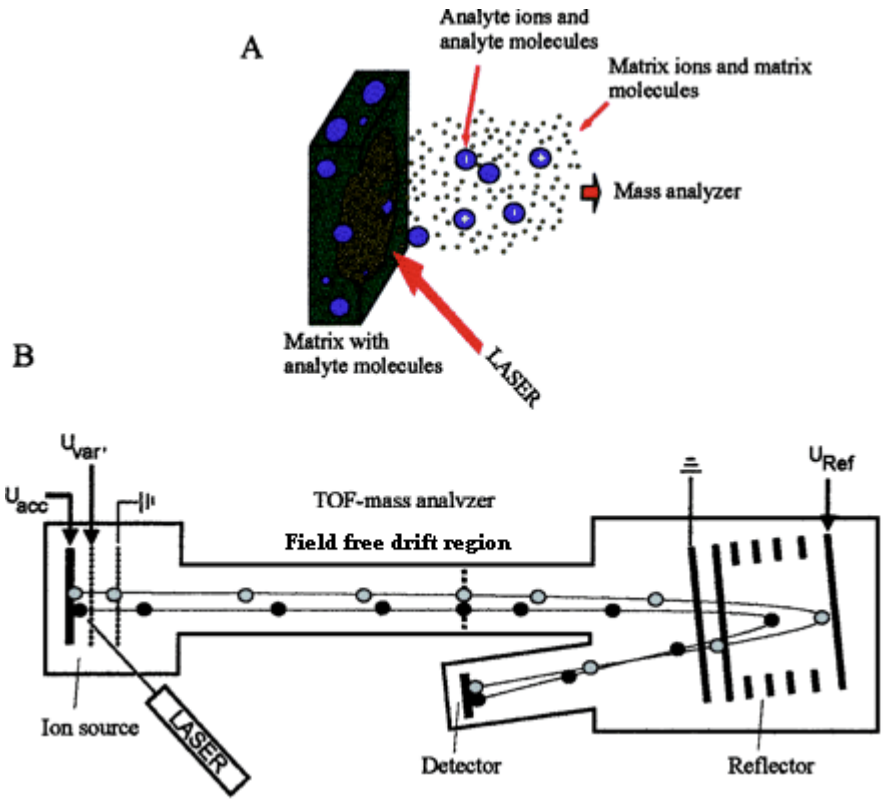


Figure 1-16: A simplified MALDI–TOF MS scheme. A: Matrix Assisted Laser Desorption Ionisation (MALDI), a laser ionizes the matrix with analyte molecules and lead them into to gas phase. B: Time of Flight (TOF) analyser, which allows determination of the mass and charge of the ions. (http://www.biochem.arizona.edu/classes/bioc471/pages/Lecture23/Lecture23.html)

Samples analysed by MALDI are first co-crystallized with the matrix. A UV-laser ionizes the mix of matrix and peptides and leads it into a gas phase (Figure 1-16 A). In MALDI-TOF, MALDI is connected to a Time of Flight (TOF) tube for mass analysis. When the ions reach the gas phase in the TOF tube they are accelerated in an electric field and separated in a field-free tube (Figure 1-16 B). The time it takes for an ion to drift trough the field-free tube (Time of Flight) depend on the charge of the ion (z) and its mass (m) (Manz 2004). The separation of the ions is detected and a data system produces a spectrum showing the ions sorted by mass (m/z) and intensity (Figure 1-17). The tops exceeding a certain intensity value are selected by a computer program and listed. These values are used for search in a database, for example Mascot.

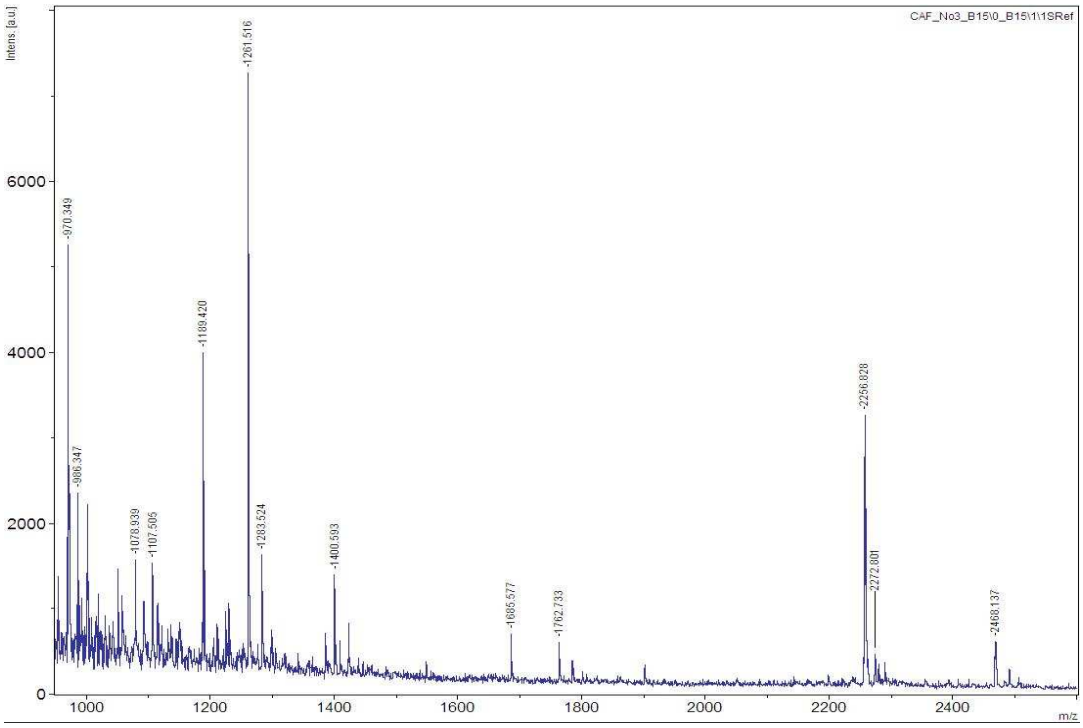


Figure 1-17: MALDI-TOF MS spectrum. The tops represent peptides at sorted by m/z (x-axis) and idensity (y-axis).

DeNovo sequencing is a different technique where the spectrometer gives the user the amino acid sequence of a small fragment usually consisting of 5 to 20 aminoacids. The protein can then be identified by search in protein sequence databases using protein–blast (blastp).

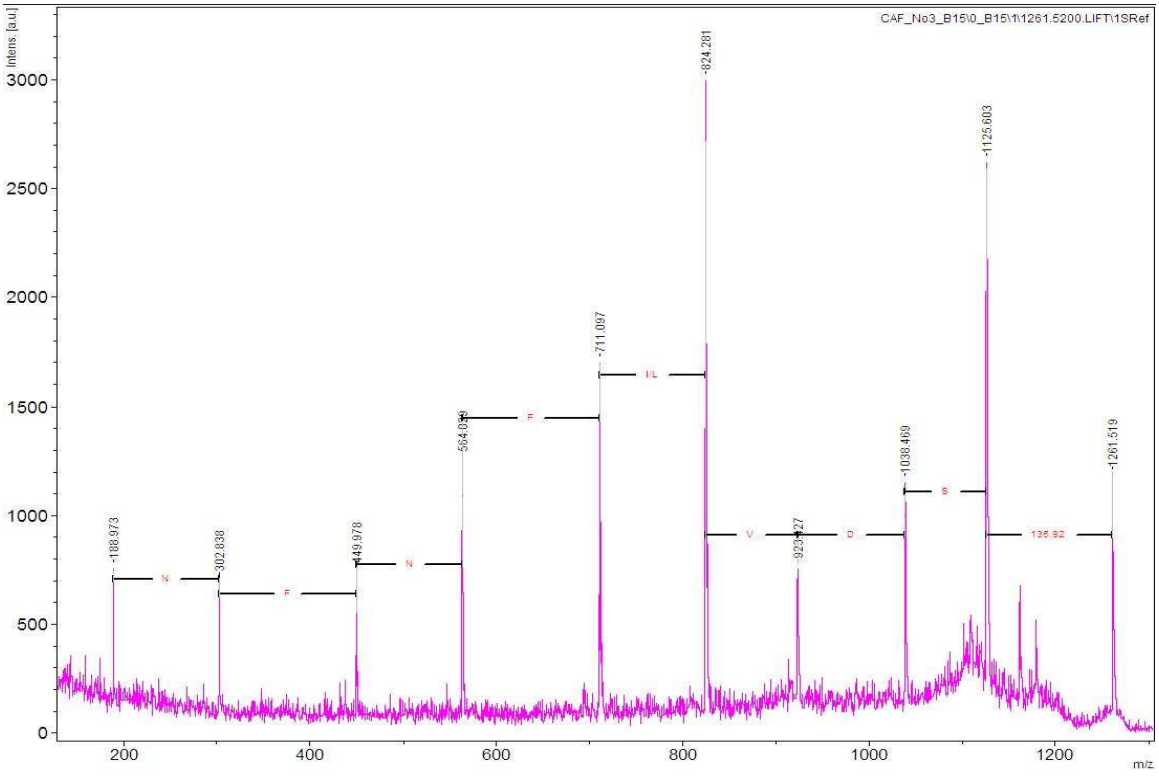


Figure 1-18: An example of a TOF/TOF sectrum by CAF fragment 1261,5Da by m/z giving the sequence S D V I/L F N F N.

A preCAF (Chemically Assisted Fragmentation) spectra of the tryptic peptides are first obtained by MALDI-TOF MS. This is compared with a post-CAF spectrum of the tryptic peptides which has been modified by derivatization. The derivatization reaction in deNovo sequencing has two main steps: Conversion of lysine side chains to homoarginine (adds 42Da to each Lysine residue) and introduction of a sulfonic acid group at the N-terminus of the peptides (adds 136Da to each peptide). A Post Sourve Decay (PSD) spectrum is produced of the selected precursor mass by using a MALDI-TOF/TOF MS (tandem Mass Spectrometer) instrument (Figure 1-18). PSD refers to a technique were the sample is fragmented before it reaches the field-free drift region. The fragments are sorted by the computer program by the m/z relationship and give the user the aminoacid sequence of the fragment. The complete protocol for PMF and deNovo sequencing is in found in Appendix G.

1.11 Lipopolysaccharide (LPS)

.

LPS is a vital part of the outer membrane of Gram- negative bacteria and is a major virulence factor. It is composed of three components: Lipid A, a polysaccharide core and the O-antigen (Figure 1-19). The lipid A is anchoring LPS to the outer membrane and is made out of fatty acids and sugar. The polysaccharide core chain is linked with lipid A and divided into a conserved inner core and a slightly variable outer core region. The O- antigen is a hypervariable polysaccharide chain often made out of peculiar sugars and with many repetitive oligosaccharide units (O-units). The diverse combinations of O-units and variation of chemical linkages are responsible for the diversity of the O-antigen and makes it possible to divide bacteria into separate serotypes (Coimbra et al. 2000).

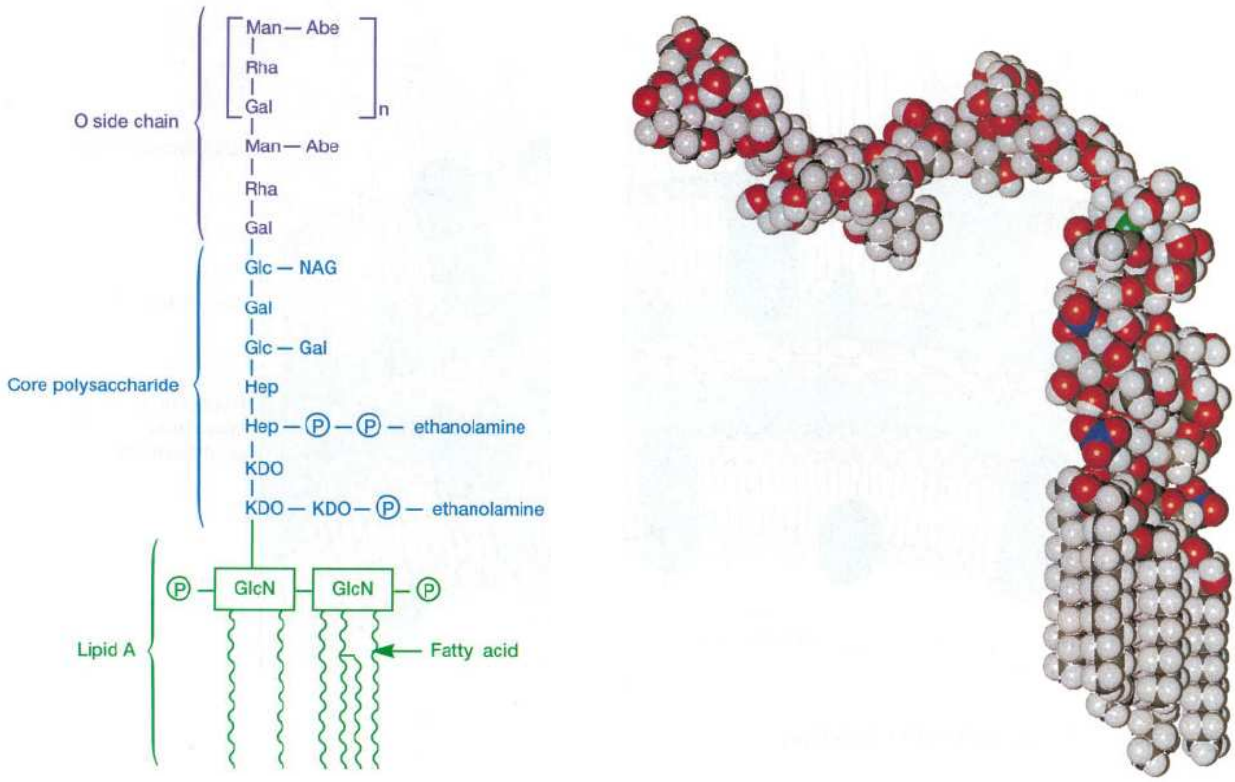


Figure 1-19: The lipopolysaccharide from *Salmonella* illustrated by a simplified diagram of on form of LPS (left) and a molecular model from *E. coli* (right). The lipid A core polysaccharides are straight and the O-side chain is bent at an angle in this model (Prescott 2005).

Together with membrane proteins lipoproteins can be extracted, separated on a gel, and visualized with coomassie staining and submitted to immunodetection by Western blotting. Some LPS might be included in the membrane protein extracs but will not be visualized by coomassie staining. The earlier methods for LPS extraction included hot phenol/water were complex, time consuming and not always reliable. Only one LPS extraction kit is available at the global market. This kit (made by iNtRON) is more effective and allows rapid and convenient extraction which can be confirmed by SDS-PAGE and silver staining.

2 Materials and Methods

2.1 Strains

– Iso10

An environmental strain isolated from a lake outside Dhaka city in Bangladesh. Cross- reacts efficiently with *Shigella* specific antisera, but does not possess *Shigella* virulence genes.

– *Shigella boydii* 15, strain ATCC12034:

Virulent type strain compared with Iso10 in proteomic and immunologic experiments.

– *Escherichia fergusonii*, strain ATCC 35469: Reference strain

– *Escherichia coli* K 12, DSMZ498: Reference strain

– *Pseudomonas aergunosa*, NCIB 8295: Reference strain

2.2 PCR Analysis

2.2.1 Cultivation and DNA isolation

Chromosomal DNA was prepared following the recommended protocol for "GenElute ™ Bacterial genomic DNA kit" (SIGMA) where DNA is isolated by using a silicia-based membrane to bind DNA.

Strain Iso10 was grown in LB medium overnight at 37°C in a shaking incubator at 200rpm. An overnight LB broth culture was pelleted and incubated at 55°C in a chaotropic salt–containing solution. DNA binds to the silica-based membrane and is separated from the lysate solution by centrifugation. The column is washed before the DNA is eluated into an ependorf tube.

2.2.2 PCR amplification of housekeeping genes

The PCR reaction is an important technique in biotechnology that makes it possible to produce thousands to millions of copies of one specific gene or gene fragment. With two oligonucleotide primers (forward and reverse) framing the gene, DNA polymerase synthesises new DNA strands from the free 3'OH end of each primer (Mullis et al. 1986).

To carry out this reaction DNA polymerase needs a proper reaction buffer, dNTPs and $MgSO_4$ at the correct concentration (Table 2–1). The amplification is regulated by temperature changes in three different stages:

1. Denaturation(96–98°C): The hydrogen bonds between the complementary bases are broken and double-stranded DNA (dsDNA) is denatured into single-stranded DNA (ssDNA).

2. Annealing: (40–60°C): Primers hybridise to the complementary sequence on ssDNA. The annealing temperature depends on the Tm of the primer set used in the PCR reaction. Length and GC% of the primer sequence influence the optimal annealing temperature.
3. Extension: (72°C): The polymerase synthesises a new strand from the free 3'OH group on the primer by incorporation of successive complementary nucleotides to the chain.

Table 2-1 Reaction mix for PCR amplification with Dynazyme polymerase

Components	**Volume added (µl)**
Sd H2O	35.5
10 x PCR reactionbuffer	5
dNTP (10mM of each)	1
$MgSO_4$ (50mM)	2
Primer F (25pmol/µl)	2
Primer R (25pmol/µl)	2
Polymerase, Dynazyme (2units/lµl)	0.5
DNA, template (25ng/µl)	2
Total volume	**50**

Table 2-2 Primers used in this study

Gene amplified	**Primer sequence**	**Product length**	**Annealing temperature**
adk	F 5'-ATTCTGCTTGGCGCTCCGGG-3' R 5'-CCGTCAACTTTCGCGTATTT-3'	583bp	54°C
fumC	F 5'-TCACAGGTCGCCAGCGCTTC-3' R 5'-GTACGCAGCGAAAAAGATTC-3'	806bp	54°C
gyrB	F 5'-TCGGCGACACGGATGACGGC-3' R 5'-GTCCATGTAGGCGTTCAGGG-3'	911bp	60°C
icd	F 5'-ATGGAAAGTAAAGTAGTTGTTCCGGCACA-3' R 5'-GGACGCAGCAGGATCTGTT-3'	878bp	54°C
mdh	F 5'-ATGAAAGTCGCAGTCCTCGGCGCTGCTGGCGG-3' R 5'-TTAACGAACTCCTGCCCCAGAGCGATATCTTTCTT-3'	932bp	60°C
mdhRmod	R 5'-TTAACGAACTCTTGCCCCAGAGCGATATCTTTCTT-3'	932bp	60°C
mdh 2008	F 5'-ATGAAAGTCGCAGTCCTC-3' R 5'-GGCGATATCTTTCTTCAGCG-3'	932bp	50°C
mdh I F	F 5'-IIIIIIIIATGAAAGTCGCAGTCCTC-3'	932bp	55°C
Ny2mdhR	R 5'- GGGTATCTGTCCCAACGTCGTGGCGG-3'	932bp	60°C
recA	F 5'-CGCATTCGCTTTACCCTGACC-3'	780bp	58°C

	R1 5'-AGCGTGAAGGTAAAACCTGTG-3'		
purA	F 5'-CGCGCTGATGAAAGAGATGA-3' R 5'-CATACGGTAAGCCACGCAGA-3'	816bp	54°C
16S rRNA	F 5'-GAGTTTGATCCTGGCTCAG-3' R 5'-GAAAGGAGGTGATCCAGCC-3'	1400bp	63°C
trpB	F 5'-ACAGAGATTCGTAGCGTT-3' R 5'-GAAAGGAACAATGACAAC-3'	1174bp	50°C
purN	F 5'-AATGCCAACGGTGAAAACGC-3' R5'-GGAGCCTTGTTCTTCGCTAA-3'	1074bp	53°C

The amplification reaction was performed on a Bio-Rad "DNA Engine Peltier Thermal Cycler" in 200 µl PCR– tubes with the following program:
96°C x 3min (96°C x 30sec – annealing temperature x 30sec – 70°C x 30sec) x 30, 70°C x 10 min – 4°C ∞
An alternative reaction mix and PCR–program was used for Phusion polymerase (Appendix D). To be able to analyse the result by using *Escherichia coli* MLST Database managed by "Max-Planck Institut für Infektionsbiologie" the following housekeeping genes were sequenced: *adk, fumC, gyrB, icd, mdh, purA* and *recA.* A new sequencing of the complete 16S rRNA gene was also carried out. Two additional genes were also sequenced (*purN* og *trpB*) (Lan et al. 2004), but they were not used in the MLST analysis.

2.2.3 Agarose gel electrophoresis

1% agarose gels in TAE– Buffer is used to visualize the PCR product. It shows the quantity (strong or weak bands), length (compared to 1Kb+ standard) and specificity (number of bands) of the PCR product.

Agarose (SeaKem LE) was dissolved in 1 x TAE buffer (40mM Tris acetate, 1mM EDTA) and melted in a microwave oven. After cooling to approximately 60°C the agarose was poured into a gel tank and a comb was inserted. The gel was allowed to cool down for approximately 30 min before the comb was removed and the gel was covered in TAE- buffer and ready for electrophoresis.
5µl of each sample was mixed with 1µl of 6 x loading buffer (0.25%, w/v) bromphenol blue, 30% (v/v) glycerol) on parafilm and loaded into separate wells.
3 µl of a 1Kb+ DNA ladder was loaded into the first well to be able to estimate the length of the PCR– product. The gel was run at 90V for 30– 45min.

2.2.4 Visualisation of PCR products with Etidium Bromide (EtBr)

Ethidium ions fluoresce when exposed to ultraviolet light. Since ethidium intercalate between basepares in double stranded DNA, EtBr can be used to visualize DNA in agarose gels.

The gel was stained in a EtBr solution for 15– 30 min followed by destaining in tap water for 15–30min. The DNA was visualized by the UV Transilluminator CHEMI Genius imaging system using Genesnap software from Syngene.

2.2.5 Purification of PCR product

After the PCR reaction the product must be separated from other components that may disturb the sequencing reaction (remaining nucleotides, primers, polymerase etc.). This was done by following the recommended protocol the "Genelute™ PCR clean–up kit" (SIGMA). The PCR product binds to a silica-based membrane in a spin column. The column is washed before the clean PCR product is eluted into an Eppendorf tupe.
Alternatively the PCR product can be purified from a 0,8% Agarose gel in TAE–buffer by following the protocol for "Genelute™ Gel extraction kit" (SIGMA).

2.2.6 Quantification of DNA

To measure the quantity of DNA two methods were used:

1. Spectrophotometric quantification at 260 nm using a Cary™ 100ConC UV–visible spectrophotometer (Varian).

   DNA absorbs light at 260nm. One A260 unit of dsDNA is equivalent to 50 µg/ml.
   The following formula was used to calculate the amount of DNA:
   DNA (µg/ml) in sample = A 260 x 50µg/ml x dilution factor

2. NanoDrop (ND100 spectrophotometer)

   1µl of sample was loaded and the instrument automatically calculates the DNA concentration in ng/µl. No need for dilutions and cuvettes.

2.2.7 Sequencing of PCR products

The most frequently used method to sequence PCR–products is the Chain Terminator Method, also referred to as the dideoxy method or the Sanger method (Sanger et al. 1992).

The principle for this method is quite similar to the Polymerase Chain Reaction except that only one primer is used and in additions to the four different deoxynucleotides (dNTPs), dideoxynucleotides are added to the reaction mix (ddATP, ddGTP, ddCTP, ddTTP). Dideoxynucleotides are "terminator nucleotides" marked with different fluorescent chemical groups. When a ddNTP is added to the growing chain the reaction terminates because there is no free 3'OH hydroxyl group for incorporation of the next nucleotide. Eventually the reaction mix will contain DNA–chains of different sizes from one nucleotide to hundreds of nucleotides all terminated by different ddNTPs. The DNA chains of different sizes are separated on a gel and a fluorescence detector register the nucleotide sequence.

The components in the sequencing reaction were mixed as shown in table 2–3. For every gene the forward and the reverse primer were used in two different sequencing reactions to achieve overlapping sequences. The synthesis of DNA fragments was performed in a Robocycler gradient 96 (Stratagene) using the following program:
96°C x 3min (96°C x 0.3min – 50°C x 0.3min – 60°C x 4min) x 65 – 6°C ∞.
The samples were delivered to the Sequencing Center at the High Technology Center in Bergen (HIB) where the sequence analysis was performed by an ABI PRISM® 3700 DNA analyzer (Perkin Elmer).

Table 2-3 Reaction mix for sequencing

Components	**Volume added (µl)**
sdH_2O	**x**
Primer (3.2µM)	1
Sequencing buffer	1
Big Dye version 3.1.	1
Template (200 ng)	x
Total	**10**

2.2.8 Phylogenetic analysis

The forward and reverse sequences were aligned by using the REVSEQ and MERGER software as implemented in the European Molecular Biology Open Software Suite (EMBOSS) available online (http://proteas.uio.no/emboss).

The online databases WU-Blast by EMBL-EBI (http://www.ebi.ac.uk/Tools/blast2/nucleotide.html) and blastn by NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi) were used in the first nucleotide sequence searches to identify the closest related species to Iso10.

The multiple construction of 22 concatenated sequences in FASTA– format were aligned using the program CLUSTALX 2.0.9 (Larkin et al. 2007).
The phylogenetic relationship was analysed by using evolutionary distances and maximum–likelihood methods with the phylogenetic inferece package PHYLIP 3.6. (http://evolution.genetics.washington.edu/phylip.html).
Distances between pairs of sequences were estimated by using the DNADIST program followed by tree construction with neighbor-joining algorithm with the Kimura two-parameter correction based on 3363 positions. Bootstrap percentage values for each branchpoint of the tree were calculated using SEQBOOT. The tree was plotted using the DRAWTREE program. *Salmonella enterica* was defined as an outgroup.

2.3 Protein extraction and fractionation

All buffers and solutions used for protein extraction and fractionation are described in Appendix B

2.3.1 Bacterial growth conditions

Cultures of Iso10 and *E. fergusonii* ATCC 35469 were grown in the General Microbiology laboratory at Jahnebakken 5, Bergen University. Cultures of *S. boydii* ATCC12034 were grown at the Microbiology and Immunology Department at Haukeland Hospital, Bergen. Cultures were grown in 100ml Luria Bertani broth at 37°C overnight in a shaking incubator (200rpm). Overnight cultures were diluted 1:10 (to a total volume of 1 liter) and grown until an OD_{600} of 0.5 was reached which ensures that both cultures were in the same growth phase.

Cultures were pelleted at 7500g, at 4°C for 10 minutes. The pellet was washed in washing buffer and pelleted again.

2.3.2 Lysis of bacterial cells

To be able to analyse the proteins from bacteria the cells have to be disrupted. This can be done by two different methods: Sonication or French Press. Generally French Press is a faster and more effective method, but can only be used for volumes larger than 5 ml. In this study,

sonication was used most frequently to disrupt the cells because this makes it easer to perform under aseptic conditions and reduces the risk for contamination of the surroundings with pathogenic bacteria.

2.3.2.1 Sonication

The principle for sonication is to apply ultrasonic sound energy to disrupt the bacterial cells.

The cells were sonicated on ice at an output control of 2– 4 with 30% duty cycle on a Sonifer II W-250 (Brandson) sonicator.
The disruption of the cells was confirmed by microscopy

2.3.2.2 French Press

The principle for the French Press is that the cells are disrupted due to rapid change of pressure. The sample is placed in a compartment of high pressure. A valve can be opened and the pressure is rapidly decreased. The sudden change in pressure disrupts the bacterial cells effectively.

The sample was disrupted twice in the French Press (Thermo electron coporation) at 1200 psi. The disruption of the cells was confirmed by microscopy.

2.3.3 Soluble protein fractionation

Two techniques were used to extract soluble proteins in this study.
The first method, described by Jennison et al., avoids precipitation techniques and should therefore give an accurate and complete representation of the soluble proteins (Jennison et al. 2006). The second method includes standard protein precipitation with TCA in acetone.

2.3.3.1 Extraction of soluble protein without precipitation

The pellet from 1 liter bacterial culture in LB broth was washed and resuspended in 10ml lysis buffer. The sample was sonicated or lysed by French Press. Unbroken cells and cellular debris were removed by centrifugation at 30000g for 1 hour at 8°C. The supernatant was collected and stored at –80°C.

2.3.3.2 Protein precipitation with TCA in acetone

The pellet from 1 liter bacterial culture was washed in washing buffer and resuspended in 2ml 20mM Tris (pH4.7) with 0.5mM PMSF before disruption by sonication or in a French Press. Unbroken cells and debris were removed by centrifugation at 6,000 x g for 10min.1µl DNase (Promega) and 1µl RNase (Sigma) were added.
The precipitation reaction was carried out by addition of 1.5ml ice-cold 10%TCA in acetone with 20mM DTT, and the mixture was incubated for 2 hours at –20°C.
The proteins were pelleted by centrifugation at 13,000 rpm at 4°C in a Heraeus Biofuge 13 3637 centrifuge with a Heraeus Sepatech 3743 rotor and washed twice with acetone with 20mM DTT. Residual acetone was removed by air drying in room temperature. The pellet was resuspended in 400µl rehydrationbuffer and shaken for 2 hours. The proteins were stored at –80°C.
To reduce streaking effects in the gels, the proteins were cleaned using 2D clean up kit from GE healthcare prior to electrophoresis.

2.3.4 Membrane protein fractionation

The method for isolation of membrane proteins with sodium carbonate was first described by Fujiki et al.(1982). They observed by electron microscopy that the carbonate-treatment turned closed membrane vesicles into flat sheets of membranes and released peripheral membrane proteins in soluble form. In this study a modified version of the method presented by Molloy et al (2001) was used.

The washed pellet from 1liter bacterial culture was resuspended in 50mM Tris-HCl pH 7.3 with 10µl DNase to a total volume of 6 ml. The cells were disrupted by sonication or in a French Press. Unbroken cells were removed by centrifugation at 6000g for 10min and the supernatant was collected.
The supernatant was diluted with ice-cold 0.1M sodium carbonate, pH 11, to a total volume of 60ml. The carbonate-treated proteins were ultracentrifuged at an average of 77,000g at 4°C for 1.5 hour using a Beckman SW41 rotor and an Optima ^{TM} L-90 centrifuge (Beckman).
The pellet was resuspended in 50mM Tris/HCl (pH 7.3) and centrifuged again at an average of 77,000 g at 4°C for 40min. The pellet was resuspended in 1-1.5 ml rehydration buffer and

stored at -80. Membrane solubilization solution 1 and 2 were also used as alternative solutions for resuspention.

2.3.5 Protein measurement

The protein concentrations in this study were measured by using the Bradford assay (Bradford 1976). The principle for this method is that when Coomassie Brilliant Blue (CBB G250) binds to the basic amino acids of the proteins it changes the absorbance maximum from 465nm to 595nm. A linear correlation between the absorbance and quantity of a standard protein makes it possible to determine the protein concentration.

Dilutions of standard protein, Bovine Serum Albumin (BSA) with a total volume of 1ml was mixed with 2 ml of the coomassie solution in a plastic cuvette. The samples were incubated for 10min in room temperature to ensure binding. The absorbance (A595) was measured by a Carry ^{TM} 100 ConC UV–visible spectrophotometer with a blank sample made of 1ml sdH_2O and 2ml coomassie solution.

2.4 Two dimentional electrophoresis

2.4.1 Sample application and rehydration

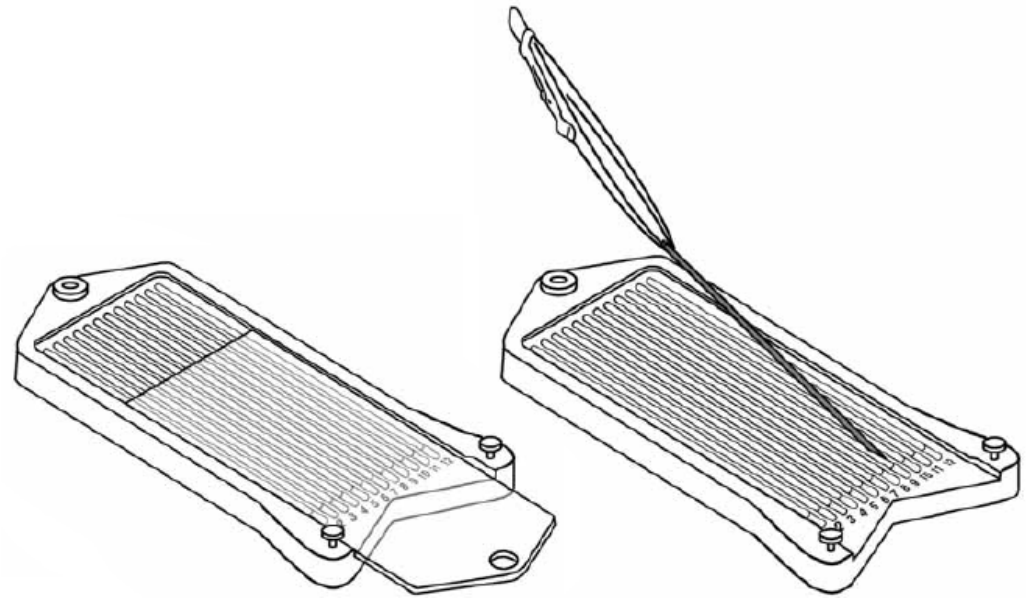


Figure 2-1: The positioning of the Immobiline Drystrip gel into the Immobiline Drystrip Reswelling Tray (GEHealthcare 2004).

The Immobilized pH gradient (IPG) strips are stored dry at –20°C and require 10–12 hours rehydration in a rehydration buffer to regain its original thickness. Rehydration of the IPG–strip can be performed together with the proteins (passive rehydration) or under low voltage (active rehydration). In this study passive rehydration was used.

The correct amount of protein diluted with rehydration buffer was adjusted to a total volume of 360 µl containing 1µg protein. The samples were applied slowly into separate slots in the Immobiline DryStrip Reswelling Tray (GE-Healthcare) (Figure 2-1). Large air bubbles were removed.

The protective cover from the 18 cm, pH 4–7 Immobiline Drystrip (GE–Healthcare) was removed and the IPG strip was applied with the gelside down onto the sample solution. Each IPG strip was covered with 3 ml of Immobiline DryStrip Cover Fluid (Amersham Biosciences) to avoid evaporation and urea crystallization. The strips were rehydrated overnight at room temperature (minimum 10 hours).

2.4.2 1st Dimension – Isoelectric Focusing (IEF)

Isoelectric focusing (IEF) is an electrophoretic technique used to separate proteins according to their differences in electric charge.

Figure 2-2: Positioning of the IPG–strips in the Drystrip aligner and connection of the electrodes (GEHealthcare 2004).

The first–dimension electrophoresis was performed with a flatbed system using Multiphor II Electrophoresis System with Immobiline Drystrip Kit (Figure 2-2).

The cooling plate was connected with the cooling tubes and the temperature was set to 20 °C. 5 ml Immobiline Dry Strip Cover Fluid was pipetted onto the cooling plate of the MultiphorII unit. The immobiline DryStrip tray was positioned on the cooling tray and 10 ml IPG Cover Fluid was poured into the tray. The DryStrip aligner was placed in the tray on top of the IPG–cover fluid.

The IPG– strips were removed from the Reswelling Tray with a pipette and blotted carefully between filter paper soaked in dH_2O. They were immediately transferred to the DryStrip aligner and placed with the gel–side up and the acidic end (+) near the anode (red electrode). When all the strips were lined up, two damped IEF– electrode strips were placed across the cathodic and anodic ends of IPG– strips. Each electrode strip must have contact with gel surface of each IPG– strip.

The electrodes were positioned on top of the IEF– electrode strips. The red electrode was aligned over the anodic end and the black electrode over the cathodic end.

The strips were covered with 80 ml IPG– cover fluid.

The strips were focused with the following program:
500V – 1 mA – 5W – 1 hour
3500V– 1mA – 5W– 16 hours

Alternative program:
500V – 1 mA – 5W – 4 hour
3500V– 1mA – 5W– 15 hours

The strips were stored at – 20 °C in Petri dishes.

2.4.3 2nd Dimention – SDS– PAGE

SDS– Page is a separation technique for proteins based on difference in molecular weight. The proteins are separated in a polyacrylamid gel containing the anionic detergent sodium dodecyl sulfate (SDS). This detergent incorporates in the proteins with a constant ratio of 1.4g SDS/g protein equivalent to one SDS molecule for every two aminoacid residues. This gives the proteins a large net negative charge proportional to their molecular weight. In addition, SDS–protein complexes assume a rod– like shape so the proteins separation in polyacrylamid gels only depend on the molecular weight of the proteins and conformational differences have no or minimal effect.

Before the proteins can be separated by SDS–PAGE the IPG–strip has to be treated with SDS–equilibration buffer with DTT followed by equilibration buffer with iodoacetamide. The equilibration buffer consists of: Tris–HCl, urea, glycerol, SDS and bromphenol blue, and has several functions. The main function is to saturate the IPG– strip with SDS which is required for the SDS– PAGE separation. Additional urea and glycerol improves the protein transfer from first to second dimension gel (Görg et al 2000) and Tris–HCl buffer stabilizes the pH gradient. The strips are first equilibrated with DTT. This secures that the proteins are completely denaturated and prevent disulfide bonds. Iodoacetamid alkylates excessive DTT to prevent point streaking (aberrant gel migration) and thiol– group reoxidation during electrophoresis.

Table 2-4 Components in the Acrylamide gel for second–dimension electrophoresis

Components	Volume
dH_2O	31.8ml
1.5M Tris–HCl pH 8.8	25ml
10% SDS	1ml
30% Acylamide	41.7ml
APS	500µl
TEMED	50µl
Final Volume	100ml

Equilibration:

After isoelctric focusing the strips are shaken in a Petri dish with equilibration buffer with DTT for 15 min. The buffer is poured off and replaced with equilibration buffer with iodoacetamid and shaken for an additional 15 minutes.

SDS–page

The glass plates were washed with ethanol and assembled. The components of the acrylamide gel were mixed (Table 2–4). Polymerisation starts when APS and TEMED are added. The gel solution was pippetted into the space between the glasses until there was 0.5cm left at the top. Butanol was pipetted on top of the gel surface to insulate the gel from oxygen and to create a flat gel surface The gel was left to polymerise for at least 1 hour and butanol was washed away. The gel can be stored at 4°C over night, but butanol should be replaced with 1x running buffer.

15µl of Mark 12^{TM} unstained protein standard was placed on a small filter paper an air–dried. The equilibrated strips are placed on the edge of dH_2O damped filter paper to remove the equilibration buffer. After a few minutes the strips were dipped in 1x running buffer and placed vertically on top of the acrylamide gel. The filter paper with molecular weight standard was placed on the gel beside the strip.

Warm 0.5% agarose (65°C) was immediately pippeted on top of the gel and left to solidify. The components of the Protean II xi electrophoresis cell unit (BioRad) were assembled and 1 x running buffer was poured on top of the gel.

The gel was run at 48mA for 6 hours (or until the blue line had reached the end of the gel) using an EPS 3501 XL power supply.

Gel Staining

1 hour in Fixation solution

1 hour in Soultion 1

3–4 days in staining solution

Gels were stored in 5% Sodium acid.

2.5 SDS–PAGE

Classic SDS–PAGE was used in this study as an alternative method to visualize membrane proteins. By using this method the proteins are only separated by molecular size. Isoelectric focusing is easily disturbed by low molecular weight ionic impurities. This can result in large regions with horizontal streaking were it is impossible to detect any spots. Classic SDS–PAGE represents an alternative method to visualize the proteins without using isoelectric focusing.

Before the samples are applied to the gel the samples are placed in a heating block (103°C) for 5 min with SDS–PAGE sample buffer with ß–mercaptoetanol. This leads to unfolding of the secondary and tertiary structure of the proteins while SDS binds efficiently to the proteins and mercaptoetanol break disulphide bridges.

The polyacrylamide gel in SDS–PAGE consists of a separating gel with a stacking gel on top (approximately 1 cm) (Table 2-5). The stacking gel is slightly more acidic and has larger pore size than the separating gel. This allows the proteins to migrate independently of molecular weight and create a concentrated band before they reach the separating gel. The separating gel is more basic and has a higher content of polyarcylamide which cause a smaller pore size. Small proteins will therefore migrate more rapidly through the gel than larger proteins.

Table 2-5: Components in the separation gel and the concentration gel for classic SDS–PAGE

Separation gel (12%)	
Components	**Volume**
dH_2O	3.35ml
1.5M Tris–HCl pH 8.8	2.5ml
10% SDS	100µl
30% Acylamide	4.0ml
APS	50µl
Temed	5µl
Final Volume	10ml

Concentration gel (4%)	
Components	**Volume**
dH_2O	6.1ml
1.5M Tris–HCl pH 6.8	2.5ml
10% SDS	100µl
30% Acylamide	1.3ml
APS	50µl
Temed	10µl
Final Volume	10ml

The glass plates were washed with ethanol and assembled. The separating gel was poured into the space between the glass plates and left to polymerize with butanol on top. After approximately one hour butanol was carefully removed and the top of the gel was rinsed with dH_2O. Stacking gel was poured on top of the separating gel and a comb was immediately placed into the fresh gel. The gel was left to polymerize and could be stored overnight at 4°C in a plastic bag with a wet filter paper.

Each well in the gel was loaded with maximum 20µg protein. SDS–PAGE sample buffer was mixed with ß–mercaptoetanol in the ratio 20:1. The sample (5–10µl) was then mixed with 10µl sample buffer/mercaptoetanol and denatured in a heating block at 103°C for 5 minutes. The solution was loaded into the wells of the gel with a low range (BioRad) and a high range (BioRad) SDS–PAGE standard on each side of the gel.

The gel was run for 45–60 min in a Mini Protein II gel chamber (BioRad) with TGS–electrophoresis buffer at 190V (Power supply model 1000/500 BioRad).

The gel was stained in 0.1 Coomassie (minimum 1 hour) and destained in a solution of 10% acetic acid/ 20% ethanol for 3 hours.

2.6 Identification of proteins

The identification of membrane proteins was performed as a commercial service provided by the PROBE platform (The Proteomic Unit at the University of Bergen) using two molecular techniques: DeNovo sequencing and Peptide Mass fingerprinting (PMF). Protocols for these techniques are placed in the appendix G. The theoretic background for this method is described in the introduction. Searches with deNovo protein sequences were performed using NCBI pblast (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

2.7 Western Blot

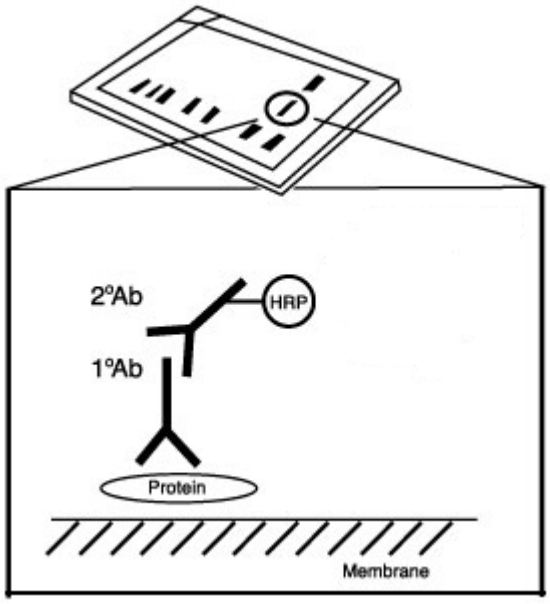


Figure 2-3: Illustration of the Western blot technique. Figure adapted from (http://www.cstj.co.jp/products/productimages/wb.jpg)

Western blot is a widely used technique to detect immunoreactive proteins and other cellular components. Primary antibodies produced by immunized rabbits bind specifically to immunoreactive antigens that have been transferred to a nitrocellulose membrane. The membrane is first incubated with primary antibodys, washed to remove unbound primary antibody before it is incubated with secondary antibody. The secondary antibody is conjugated with the enzyme Horse Radish Peroxydase (HRP). This enzyme develops a coloured product by oxidation of the substrate 4-chloro–1-naphthol in the presence of hydrogen peroxide (H_2O_2). After washing the membrane to remove unbound secondary antibodys the HRP colour buffer with the substrate is added and areas containing the antigen will appear as red spots/bands after a few seconds (Figure 2-3).

2.7.1 Western Blot of 2D gels

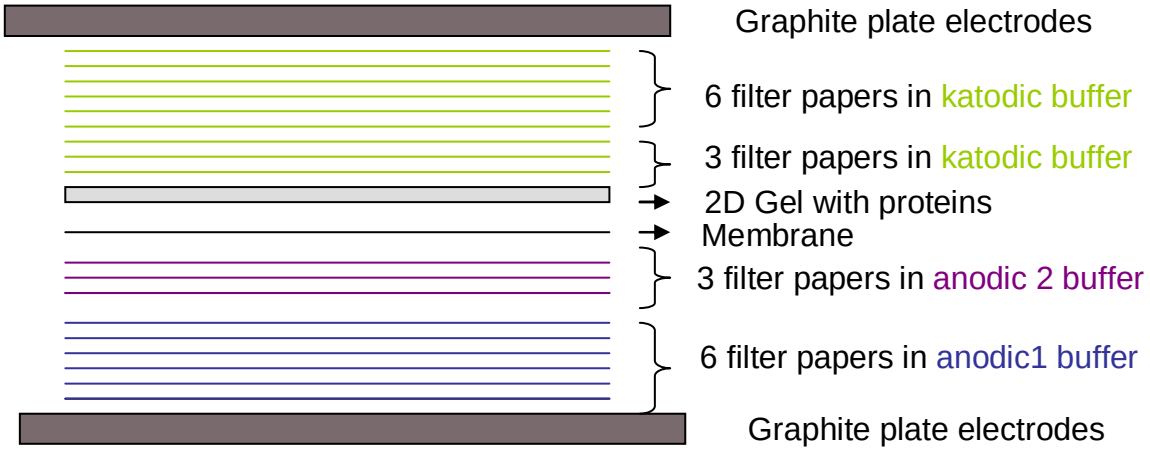


Figure 2-4: Illustration of the experimental setup of the Western blot of 2D gels.

The Western blot is performed by using a Multiphore II NovaBlot Electrophoretic Transfer Unit with two graphite plate electrodes. All buffers and solutions used for Western blotting are described in Appendix C. The *S. boydii* 15 specific antibodys used in this study was produced at the University of Dhaka as described in appendix G.

The western blotting was performed using the following steps:

1. The NovaBlot graphite plates were soaked in dH_2O for minimum 1 hour before use. The anodic graphite plate was placed into the Multiphore II and the anodes were connected.
2. Six filter papers were soaked in anodic solution 1 and placed on top of the anodic graphite plate.
3. Three filter papers were soaked in anodic solution 2 and placed on top of the other filter papers.
4. The Hybond ^{TM} N^+ nitrocellulose membrane (Amersham Biosciences) was soaked in anodic solution 2 and placed on top of the other filter papers.
5. The gel was placed on top of the membrane.
6. Three filter papers were soaked in cathodic buffer and placed on top of the gel.
7. If more than one gel were blotted, another membrane was soaked in anodic solution 2 and placed on top of the three filter papers and step 3 to step 6 were repeated.
8. Finally six filter papers were soaked in cathodic buffer and placed on top of the other filter papers.
9. The kathodic graphite plate was placed on top of the sandwich and the anodes were connected to the Multiphor II Unit.

10. The electroblotting was run at 150 mA for 1 hour and 45 min.

After each steps described above, a glass tube was rolled over the filter papers or the membrane to remove air bobbles.
The Precision Plus Protein™ All Blue standard (BioRad) was run with the protein samples on the 2D–gels and SDS–PAGE gels which were later blotted to membranes. This standard is prestained and protein bands are visible in the gel without the coomassie treatment.
If the All Blue standard was observed on the membrane the protein transfer from the gel to the membrane had most likely been successful. To be completely certain the membrane could also be stained with amido Black to visualize the protein.

2.7.2 Western blot of SDS-PAGE gels

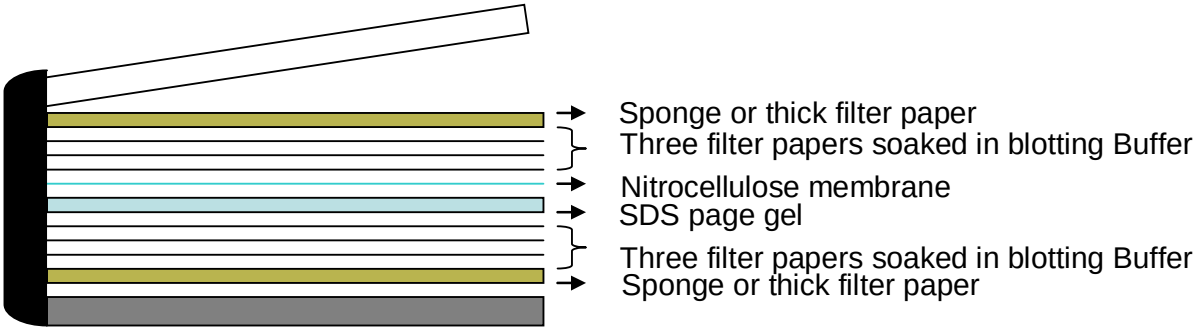


Figure 2-5: Illustration of the experimental setup of the Western blot of SDS–page gels

The western blotting of SDS- page gels was done by using the following steps:

1. The gel and the nitrocellulose membrane were incubated in 1 x Blotting buffer for 30 min.
2. The sponge and the filter papers were soaked in blotting buffer.
3. The soaked sponge is placed on top of the grey side of the chamber
4. Three saturated filter papers were placed on top of the sponge.
5. The gel was transferred from the blotting buffer and placed on top of the filter papers.
6. The nitrocellulose membrane was placed on top of the gel. A glass stick was rolled over the membrane to remove air bobbles and excess buffer.
7. Three soaked filter papers were placed on top of the membrane.
8. One sponge saturated in blotting buffer was placed on top of the membranes.
9. The “sandwich” was placed in a MiniTrans-Blot gel chamber (BioRad) together with a cooling element and filled with 1x Blotting buffer.
10. The proteins were electroblotted at 100V for 1 hour.

2.7.3 Hybridisation of immunoreactive proteins

Hybridisation of immunoreactive proteins was carried out by following the recommended protocol for Immun–Blot–Assay Kit with goat AntiRabbit lgG (H&L) and Horse Radish Peroxidase (BioRad) with a few modifications.

1) **Blocking:** The blotted membrane was blocked in 2% skim milk in TBS buffer for minimum 2 hours.
2) **Washing:** The membrane was washed for 5 min in TBS buffer at room temperature with gentle shaking followed by 5 min in Tween TBS buffer and another 5 min in TBS buffer.
3) **Primary antibody:** The membrane was covered in primary antibody diluted 1:1000 with 2% skimmed milk in TBS and incubated for 1.5 hour in room temperature with gentle shaking.
4) **Washing:** As described earlier
5) **Secondary antibody:** The membrane was covered with a solution of secondary antibody (HRP conjugated anti rabbit lgG, developed in goat, Sigma Immuochemicals) diluted 1:3000 with 2% skimmed milk in TBS for 1 hour in room temperature with gentle shaking.
6) **Washing:** As described earlier.
7) **Staining:** The membrane was calibrated for a couple of minutes in citrate buffer before it was covered in freshly made substrate buffer until the spots appeared (not more than 45 min). The membrane was then washed with dH_2O serveral times to stop the reaction.
8) The picture was taken while the membrane was still wet and the membrane was stored dark between filter papers.

2.8 LPS– Extraction

Lipopolysaccharide (LPS) is an important part of the outer membrane of gram– negative bacteria.

In this study LPS was extracted by adding chloroform to lysed bacterial cells. The chloroform will separate from the original solution and move down to the buttom of the tube while a white line is formed between the two layers. This region consists of cell debris, protein, genomic DNA and RNA. LPS is left in the upper layer and moved to a new tupe. The LPS is purified and washed with solutions provided by the kit and run on a SDS–PAGE.

The LPS extraction was done by following the recommended protocol for "LPS Extraction kit" as provided by iNtRON Biotechnology.

5 ml of an overnight culture was centrifuged for 30 sec at 13,000 rpm in a Heraeus Biofuge 13 3637 centrifuge with a Heraeus Sepatech 3743 rotor at room temperature. The supernatant was removed and 1 ml of lysis buffer was added to the pellet. The tube was vortexed carefully to dissolve the pellet properly.

After 200µl of chloroform was added the tube was vortexed and incubated at room temperature for 10 min. The tube was then centrifuged for 10 min at 4°C and 400µl of the supernatant was transferred to a new tube without touching the white precipitation.

To purify LPS from other components of the cell extract (e.g. protein, DNA, lipids) 800µl of purification buffer was added and the solution was incubated for 10min at –20°C. The tube was finally centrifuged at 13,000 rpm for 15 min at 4°C.

To remove impurities (e.g. salts), the pellet was washed with 70% ethanol and dried.

70µl of 10mM Tris– HCl (pH 8.0) was used for resuspension of the LPS and the solution was boiled for 1 min.

The LPS was run on a SDS–page and blotted on a nitrocellulose membrane by using the same technique as described earlier. ß–mercaptoetanol was not added to the sample buffer. A parallel gel was stained with Silver staining to confirm that the LPS extraction was successful. LPS from Iso10 and *Shigella boydii* was extracted at Dhaka University. LPS from *E. fergusonii* was extracted at University of Bergen.

2.9 Silver staining

The LPS run on SDS–page gels was visualized by silver staining. The silver staining was performed by following the recommended protocol for "Silver Staining kit" as provided by Invitrogen.

The following solutions were prepared immediately before staining by the reagents provided by the kit: Sensitizing solution, Staining solution and Developing solution.
After the electrophoresis the gel was rinsed in ultrapure water and incubated in fixative solution (40% ethanol, 10% acidic acid in sdH_2O) for minimum 20min. The gel could be stored in fixative solution overnight if there was not enough time to complete the staining procedure the same day as the electrophoresis.
After fixation the gel was washed in 30% ethanol (10 min) and incubated in Sensitizing solution for 10 min. To remove the Sensitizing solution the gel was first washed in 30 % ethanol and then washed in sd H_2O. After 15 min incubation in Staining solution the gel was briefly washed in sd H_2O (20–60sec). It was important not to wash the gel for more than a minute after the staining because this would remove silver ions and decrease the quality of the gel. The gel was incubated in Developing solution until the bands appeared and 10ml Stopper solution (provided by the kit) was immediately added on top of the gel. It was important that the stopper was not added too late as this would lead to a dark background.
The gel was washed in sd H_2O for 10 min and a picture was taken.

2.10 Biochemical identification with Api20E

API 20E (Biomérieux) is a standardised biochemical identification system for bacteria in the family *Enterobacteriaceae* and some other gram negative, rod shaped bacteria. The strip contains 20 microtubes with dehydrated substrates which are used to determine the metabolic capabilities of the bacteria. Each microtube is incubated overnight with bacterial suspension. During the incubation period the metabolism produce a colour change which can be easily detected the next day. An additional oxidase test can be necessary for full identification in some cases.

5ml dH_2O was poured into the tray to create a damp environment. A colony from a fresh culture was transferred to 5ml 0.85% NaCl and slightly shaken to create a homogenic suspension. A complete list of the twenty tests is shown in Table 3–5 in the results chapter.

The strip was placed in the tray and the bacterial suspension was poured into the wells with a pipette.
In the wells for trinatriumcitrat (CIT), natriumpyruvate (VP) and gelatine (GEL) both the tube and the well was covered with suspension.
In the wells for L–arginine (ADH), L–lysine (LDC), L–ornithine (ODC), natriumthiosulphate (H_2S) and urea (URE) the wells were covered with mineral oil to create an anaerobic environment.
The incubation box was closed and incubated at 37°C for 18–24 hours.
After the incubation period the strip was read and compared with the reference table.
Three of the tubes need processing by additional reagents before the result can be determined:

- L– tryptopane (TDA) test: One drop of TDA reagent was added to the well. Red colour indicates a positive reaction.
- Sodiumpyruvate (VP) test: One drop of each of the reagents VP1 and VP2 was added to the well. The result was read after 10min. A light red or red colour indicated a positive result. A white or slightly red colour indicates a negative result.
- L– tryptophane (IND) test: The test for indol production should be done at the end of the procedure since the reaction releases gases that may affect the other tests. One drop of the JAMES reagent was added to the well. If a light red colour develops in the entire well the reaction was positive.

The oxidase test was performed by transferring a colony from a fresh culture to a filter paper. A colony from an *E.coli* K12 (DSMZ 498) culture was used as a negative reference and a colony from a *Pseudomonas aergunosa* (NCIB 8295) culture was used as a positive reference. One drop of the Oxidase Reagent (BioMérieux) was added on top of the culture. A picture of the colonies was taken after 1min. Blue colour indicates a positive result.
The results can be analysed by comparing the test results with the BioMérieux identification table or by calculating a numerical profile and use the online APIweb^{TM} identification tool (https://apiweb.biomerieux.com/servlet/). To use the APIweb identification a result paper from BioMérieux (figure 3-5 in the results chapter) was filled out. The tests on the strip are separated in groups of three and a positive test reaction equals a number of 1, 2 or 4. The addition of the three values in each group gives one of the numbers in the seven- digit code. This code was inserted to BioMérieux's online database resulting in the identification of the strain.

3 Results

Iso10 was isolated outside Dhaka city in Bangladesh using protocols designed for isolation and cultivation of S*higella* bacteria. Earlier studies has shown that it effectively cross-reacts with S*higella* specific antisera but does not possess the genes involved in *Shigella* virulence. The following section presents the results of my study focusing on the genetic and proteomic relationship of Iso10 and *Shigella*.

3.1 16S rRNA gene sequencing

Since only the V3– region of the 16S rRNA gene was sequenced in earlier studies of strain Iso10, a new sequencing of the complete 16S rRNA gene was carried out (Figure 3–1). The hit with the highest score (Table 3-1), *Shigella* sp. BBDP80, had a 99.86% identity match with only 2 mismatches out of 1429bp. Still the other top five hits, *E. fergusonii*, *E. coli* and *S. Sonnei* only had four mismatches and two of the mismatches were common for all the hits including *Shigella* sp. BBDP80. The alignments of Iso10 with the sequences from *Shigella* spp. (acc.nb. DQ337525) and *E. fergusonii* (acc.nb. AF530475) are placed in Appendix F. This shows that 16S rRNA sequencing is not sufficient to successfully discrete discriminate among closely related strains of bacteria.

Table 3-1: Results from Blast search (WU–blast2, EMBL–EBI) with the 16S rRNA gene sequence from Iso10. All the hits had an E-value of 0.0.

Gene	Acc.number	Organism	Identity (bp)	Identity %
16S	DQ337525	*Shigella* sp. BBDP80	1427/1429	99.86
	AF530475	*Escherichia fergusonii* ATCC 35469	1425/1429	99.72
	CU928158	*Escherichia fergusonii* ATCC 35469	1425/1429	99.72
	CP000247	*Escherichia coli* 536	1425/1429	99.72
	Ab273732	*Shigella sonnei* GTC 781	1425/1429	99.72

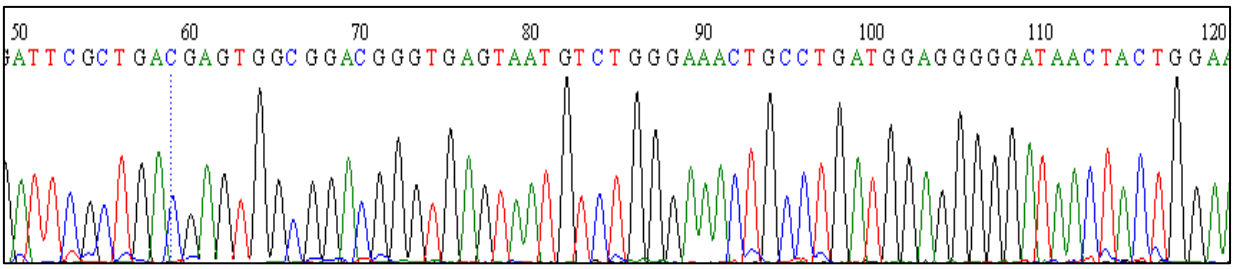


Figure 3-1: Part of chromatogram from the 16S gene sequencing of Iso10

3.2 MLST Analysis

The MLST database for *E. coli*

Nine housekeeping genes were PCR– amplified using genomic DNA from the environmental isolate Iso10. The PCR products were purified and directly sequenced by using the same primers as in the PCR amplification.

Four different reverse primers and three different forward primers were used in the attempt to amplify the malate dehydrogenase (*mdh*) gene from Iso10. After numerous attempts, a PCR–product was obtained by using the mdh 2008 forward and reverse primers designed by Pupo et al (1997) with 50°C annealing temperature. The other 9 PCR products were achieved by using the primers designed by Achtman (Wirth et al. 2006) and the trpB and purN primers designed by Lan (Lan et al. 2004).

The first Blastn searches with the sequences of the housekeeping genes gave a range of identity values (see AppendixE). The sequences coding for phosphoribosylglycinamide formyltransferase (*purN*) malate dehydrogenase (*mdh*) and fumarase (*fumC*) had poor identity match ranging from 80% to 85% identity with the corresponding genes in *E. coli.* Adenylate kinase (*adk*) and isocitrate dehydrogenase (i*cd*) had a 99% identity match with *E. fergusonii* while tryptophan synthase beta subunit (*trpB*) had a 99% identity match with *Shigella dysenteria*. The other genes, DNA gyraseB, adenylosuccinate synthetase, recombinase A (*gyrB, purA, recA*, respectively) had a 97–98% identity match with *E. coli*.

While the MLST analysis was completed seven of the sequences were edited to include only 450–550bp to be able to align them with the alleles in the Achtman database for MLST analysis of *E. coli*. Two of the genes (*trpB* and *purN*) were not included in the MLST analysis. New searches in the EMBL–EBI database with the edited, short sequences resulted in more accurate matches (Table 3–2). All housekeeping genes had 99% to 100% hit with *E. fergusonii*. The reason for this is that only short partial sequences from five of the housekeeping genes of *E. fergusonii* was available in the EBI and NCBI databases.

Table 3-2: Results from Blastn search (WU–blast2, EMBL–EBI) with partial sequences from seven housekeeping genes.

Gene	Acc.number	Organism	Identity (bp)	Identity %	E- value
adk	AY686533	*Escherichia fergusonii*	531/536	99.10	8.4e-88
	AY4464415	*Escherichia coli* strain O157:H7	503/536	93.84	1.2e-77
	AY464267	Escherichia coli strain ECOR37	502/536	93.66	1.2e-77
	AY382659	*Escherichia coli* strain 2P9	501/536	93.47	7.9e-77
	AY464188	*Escherichia coli* strain 21P	501/536	93.47	7.9e-77
fumC	AY686535	*Escherichia fergusonii*	444/446	99.55	8.5e-94
	X04065	*Escherichia coli* (EC4.2.1.2)	379/446	84.98	2.0e-66
	EA375705	Sequence 24528 from patent US 7314974.	379/446	84.98	3.2e-66
	CL669644	*Pristionchus pacificus*	379/446	84.98	5.1e-66
	AY464230	*Escherichia coli* strain ECOR21	379/446	84.98	9.5e-66
gyrB	AY370846	*Escherichia fergusonii* strain ATCC 35469	459/460	99.78	2.3e-95
	DQ386687	*E. fergusonii* strain ATCC 35469	459/460	99.78	2.3e-95
	AB083895	*Escherichia coli* strain ECO166P4.	449/460	97.61	1.1e-91
	AB083856	*Escherichia coli* strain ECO15P2.	448/460	97.39	2.9e-91
	AB083945	*Escherichia coli* strain ECO8P1.	448/60	97.39	2.9e-91
icd	AY132849	*Escherichia fergusonii* strain ATCC 35469	514/514	100	2.3e-110
	AY132845	*Escherichia coli* strain ECOR46	504/514	98.05	6.8e-106
	AY132847	*Escherichia coli* strain ECOR49	503/514	97.86	1.7e-105
	AY245921	*Escherichia coli* strain DAEC19	503/514	97.86	1.7e-105
	AY245934	*Escherichia coli* strain DAEC218	503/514	97.86	1.7e-105
mdh	EF012087	*Escherichia fergusonii*	452/452	100	5.4e-77
	AJ287796	*Escherichia coli* strain IHIT1190	392/452	86.73	1.6e-67
	AJ287797	*Escherichia coli* strain IHIT1968	392/452	87.73	1.6e-67

	AJ287795	*Escherichia coli* strain IHIT0067	390/452	86.28	1.0e-66
	AJ287798	*Escherichia coli* strain IHIT3000	390/452	86.28	1.0e-66
purA	AY686536	*Escherichia fergusonii*	477/478	99.79	2.1e-99
	CS191906	*Escherichia coli* Patent WO2005103073.	473/478	98.95	3.2e-98
	AY464270	*Escherichia coli* strain ECOR37	473/478	98.95	8.8e-98
	AY464419	*Escherichia coli* strain O157:H7	473/478	98.95	8.8e-98
	J04199	*Escherichia coli*	471/478	98.54	1.0e-97
recA	AY686537	*Escherichia fergusonii*	507/510	99.41	7.5e-106
	BD064136	Synthetic construct, artificial sequences.	501/510	98.23	5.0e-104
	BD233595	Synthetic construct, artificial sequences.	501/510	98.23	5.0e-104
	CS423861	Synth.construct. Artf. seq. Sequence 14 from Patent EP1707641.	501/510	98.23	5.0e-104
	CS443102	Synth.construct. Artf. seq. Sequence 7 from Patent EP1717322.	501/510	98.23	5.0e-104

By searching with the DNA sequence of each of the housekeeping genes in the Achtman database, an allele number representing the closest match for each gene is obtained. The Iso10 sequences gave identities from 85.4 to 100% (Table 3-3). Only the gyrB sequence matched one of the database alleles 100%. This shows that Iso10 is not represented in this database.

By inserting a profile of seven allele numbers obtained from the Iso10 sequences the database identifies the sequence types (ST) which share most allele numbers with the test organism. For Iso10, only four of the seven best-matching genes (*gyrB, icd, purA, recA*) were present in the closest related sequence types (ST770 and ST485) (Table 3-4). Iso10 is thus not closely related to any of the strains represented in the Achtman MLST database.

Table 3-3: Result from search in the *E. coli* MLST database by Mark Achtman with the DNA sequences from seven housekeeping genes from Iso10. The data for Iso10 shows the allele number for the representative genes with the highest identity match (%).

Sequence Type (ST)	Adk	fumC	gyrB	Icd	Mdh	purA	recA
Iso10	35 (93.4%)	93 (85.4%)	55 (100%)	101 (99.0%)	53 (87.5%)	40 (99.8 %)	38 (99.4 %)

Table 3-4: Result from search in the *E. coli* MLST database showing the sequence types (ST) which share allele numbers with Iso10. Allele numbers marked green are identical to allele numbers for Iso10.

Sequence Type (ST)*	Adk	fumC	gyrB	Icd	Mdh	purA	recA
ST770	52	116	55	101	113	40	38
ST485	52	116	55	101	35	40	38
ST434	52	116	55	101	35	40	78
ST747	52	54	46	48	35	40	38
ST662	120	151	46	48	35	40	38

*** All the ST matches represent only one strain except ST747 that represent three ETEC strains isolated in Egypt in the period 2003 to 2005 by the same source lab (NAMRU–3)**

The five closest related ST strains in the database represent isolates from two different continents (Europe and Africa) and three different pathogenic types of *E. coli* (EAEC, ETEC, EHEC) which causes diarrhoea in humans (Table 3-5).

Table 3-5: Supplementary information for the five sequence types (ST) closest related to Iso10 in the Achtman MLST database.

Sequence Type (ST)	Host type	Host species	Year	Country	Pathogen	Sereotype	Disease
ST 770	Food	Poultry	2004	Denmark	None		None
ST 485	Human	Human	1995	Nigeria	EAEC		Diarrhoea
ST 434	Human	Human	1995	Nigeria	EAEC		Diarrhoea
ST 747	Human	Human	2003–2005	Egypt	ETEC		Diarrhoea
ST 662	Human	Human	None	Austria	EHEC	O118:H11	None

Phylogenetic analysis

In order to elucidate the phylogenetic relationship of Iso10 and the *E. coli/Shigella* group, a phylogenetic tree based on concatenated housekeeping gene sequences was constructed (Figure 3–2). Twentyone strains from the Achtman database and from complete genome sequences in EMBL–EBI were included. The strains include isolates from eight pathogenic types: Enteroinvasive *E. coli* (EIEC), enteropathogenic *E. coli* (EPEC), enterohaemorrhagic

E. coli (EHEC), enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAggEC), extraintestinal *E. coli* (ExPEC), *E. coli* K1 and *Shigella*. Nonpathogenic *E. coli* K12, *Escherichia albertii* and *Escherichia fergusonii*, were also included in the analysis. This includes representatives for all the *E. coli* ancestry groups: A, AxB1, B1, B2, D and ABD estimated by Wirth et al (2006).

The phylogenetic tree is based on concatenated sequences of all seven housekeeping genes. 21 STs were included in a multipple alignment and a phylogenetic tree was constructed as described in methods. *Salmonella enterica* was used as an outgroup.

The phylogenetic tree clearly shows, with the support of high bootstrap values, that *E. fergusonii* and Iso10 are closely related and should be placed outside the main *Shigella-E. coli* cluster in a distinct clade and further confirms the difficulties to separate *E. coli* and *Shigella* phylogenetically.

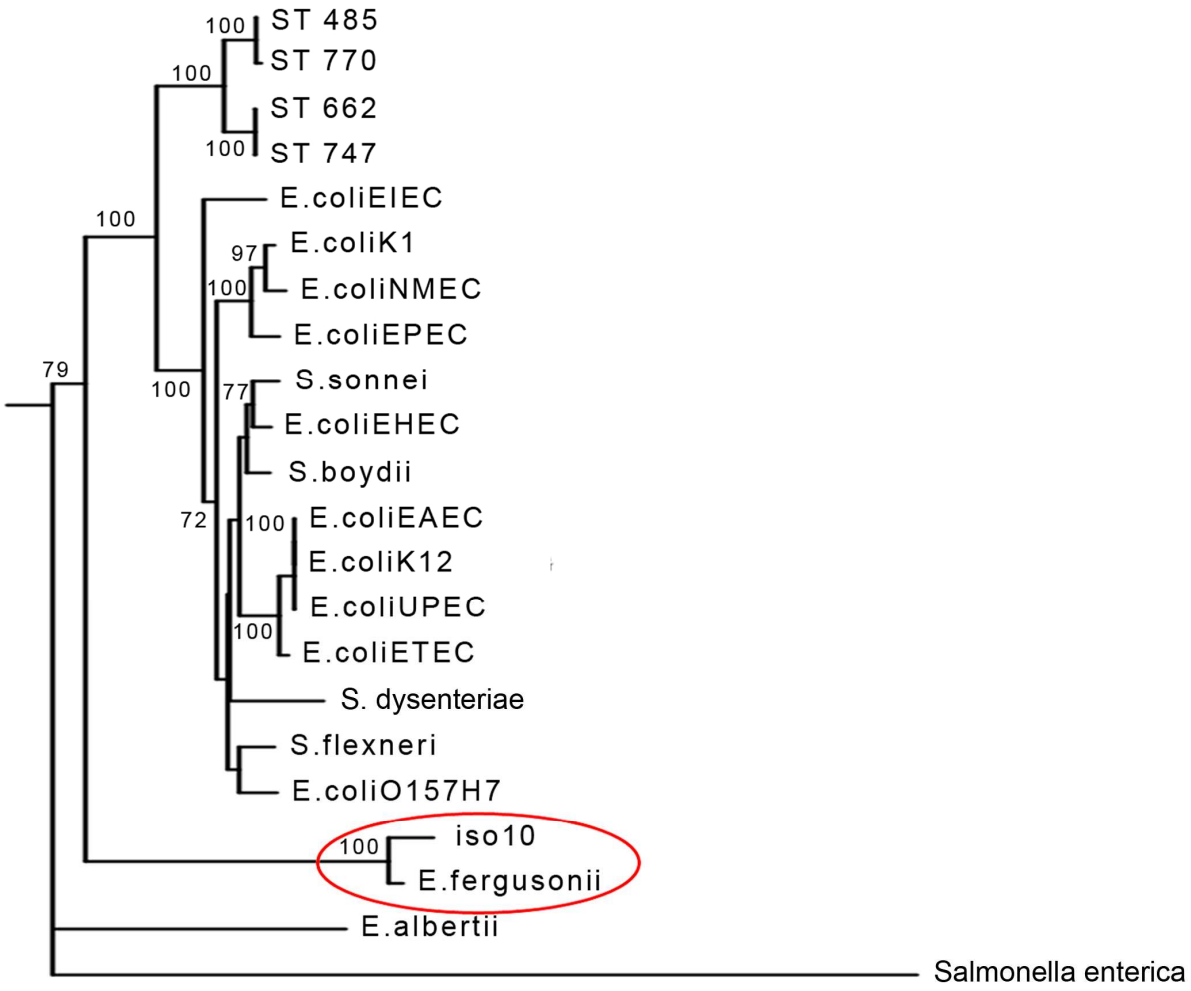


Figure 3-2: Phylogenetic relationships among 21 isolates of the *Escherichia/Shigella* group based on concatenated sequences from seven housekeeping genes. Four of the closest related sequence types (ST) from the Achtman database was included in the tree. A 50% conseous majority tree is shown with bootstraps values from 100 replicates. Only bootstraps values over 75% is shown. *Salmonella* is used as an outgroup.

3.3 Biochemical characterization with API20E

A biochemical characterisation was performed by using the API 20E kit (Figure 3–3 and 3-4). The results are presented in table 3–6 and compared with the biochemical properties of *E. coli, E. fergusonii* and *Shigella* using reference materials provided by Biomérieux.

Figure 3-3: Biochemical characteristics determined by API 20E kit for *Enterobacteriaceae.*

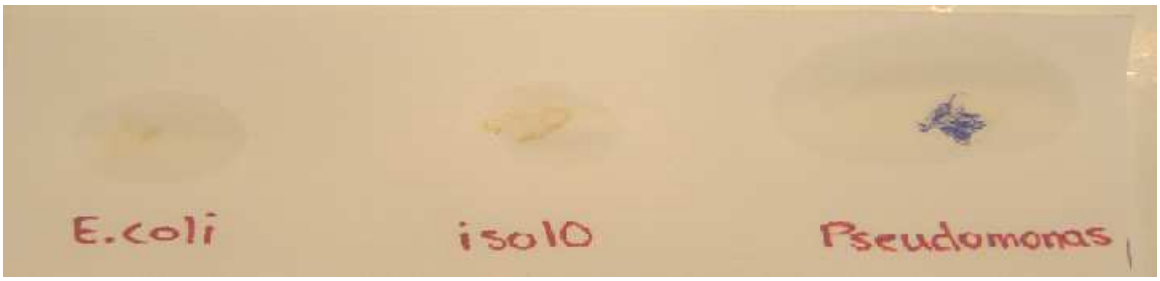


Figure 3-4:Oxidase test for Iso10 showing *E. coli* and *Pseudomonas* as negative and positive control. Blue colour indicates a positive result.

As table 3–6 shows Iso10 shares its the biochemical characteristics with *E. fergusonii* in every test except the mobility test which is positive for 93% of the *E. fergusonii* strains. It differs from S*higella* by testing positive for lysine decarboxylase, Ornithine decarboxylase and indol production and by fermenting rhamnose and amygdaline. It differs from *E. coli* by being negative for sorbitol fermentation and positive for amygdaline fermentation.
The strain can also be identified by calculating a numerical profile and using the APIweb^{TM} identification tool(https://apiweb.biomerieux.com/servlet/). The numeric profile from Iso10 was 5-1-4-4-1-1-3 (Figure 3–5). The APIweb identified Iso10 as *E. fergusonii* with 98. 9 % probability (Table 3–6). The ability to ferment L-lysine (LDC) made the second alternative, *Escherichia hermanii*, unlikely, with only 1% probability. This supports the close relationship between Iso10 and *E. fergusonii.*

Table 3-6: The Api20E results for Iso10 compared with Biomérieux's identification table for *E. fergusonii*, *E. coli* and *Shigella* presented as % positive reaction in the tested strains after 18–24 hours at 36°C ±2°C.

Reactive ingredient	Test Code	Iso10	*E. fergusonii*	*E. coli* 1	*E. coli* 2	*Shigella* spp.	*Shigella sonnei*
2– nitrofenyl ß–D–galaktopyranosid	ONPG	+	96	88	47	7	96
L–arginine	ADH	–	1	5	10	0	0
L–lysine	LDC	+	99	74	45	0	0
L– ornithine	ODC	+	100	70	30	4	97
Trinatriumcitrat	CIT	–	1	0	0	0	0
Natriumthiosulfat	H_2S	–	0	3	2	0	0
Urea	URE	–	0	2	1	0	0
L– tryptofan (Tryptophan DeAminase)	TDA		0	0	0	0	0
L–tryptofan (indol production)	IND	+	99	89	77	39	0
Natriumpyruvat	VP	–	3	0	0	0	0
Gelatine	GEL	–	0	0	0	0	0
D– glucose	GLU	+	100	99	97	96	100
D– mannitol	MAN	+	99	97	84	63	99
Inositol	INO	–	0	3	2	0	0
D– sorbitol	SOR	–	0	90	42	15	1
L–rhamnose	RHA	+	87	82	35	7	75
D– sucrose	SAC	–	0	41	4	7	2
D– melibiose	MEL	–	3	67	34	22	1
Amygdalin	AMY	+	99	20	12	0	0
L–arabinose	ARA	+	97	82	90	52	97
Oxidase test	OX	–	0	0	0	0	0
Mobility	MOB	–	93	95	5	0	0

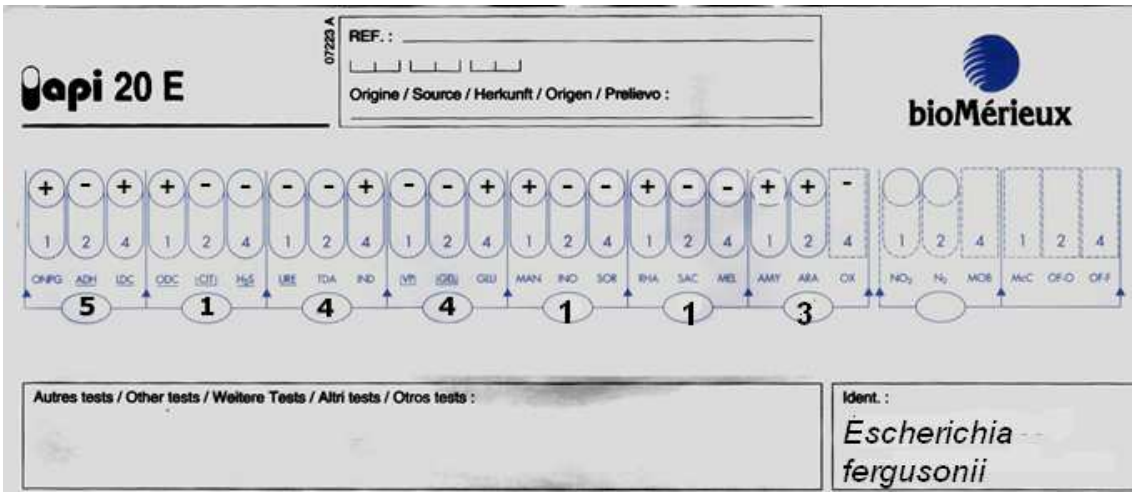


Figure 3-5: The biochemical characteristics are filled in the API 20E result paper giving the numerical profile for identification.

Table 3-7: The identification of Iso10 with numeric profile with APIweb

	Taxa	% ID	T	Tests against
1	*Escherichia fergusonii*	98.9	1.0	
2	*Escherichia hermannii*	1	0.67	LDC 1%

3.4 2D gels of soluble and membrane-associated proteins

To complement the genetic and biochemical characterization of Iso10, a proteomic analysis was performed. Solubilization and separation of membrane proteins by 2D gel electrophoresis are known to be more complex than for soluble proteins. In this study three different solutions were used to resuspend membrane proteins after ultracentrifugation: Rehydration buffer, membrane solution 1 and membrane solution 2 (Appendix B).

The rehydration buffer was effective for resuspending soluble proteins but when it was employed for resuspending membrane proteins, only a few (2-4) spots were visible on the 2D gel (picture not shown). Membrane solution 1 was more effective than the rehydration solution but caused more streaking. This solution was used to resuspend the membrane proteins from *S. boydii* shown in figure 3-7.

The best 2D gel with membrane proteins was achieved by resuspending the proteins in membrane solubilization solution 2 and by rehydrating the solution in rehydration solution. Figure 3-6 shows membrane proteins from Iso10 resuspended in membrane solution 2. Since the membrane proteins that were analysed were from an unidentified, environmental isolate, it was chosen to use deNovo sequencing as the main identification method. Some of the spots that failed to give results on deNovo sequencing were analysed by Peptide Mass Fingerprinting (PMF).

Fifty-two spots from the Iso10 membrane fractions were visualized with coomassie staining. Twenty-five of these spots were excised from the gel for identification, giving 16 positive results and identification of 10 different proteins (14 spots identified with deNevo sequencing, 2 proteins with PMF) (Table 3-8 and Table 3-9). The identified proteins represented five different outer membrane proteins (OmpA, OmpC, OmpF, OprF, OmpX), three lipoproteins (peptidoglycan associated, LppB, MetQ) and two contaminating cytoplasmic proteins (GroEL and AtpD).

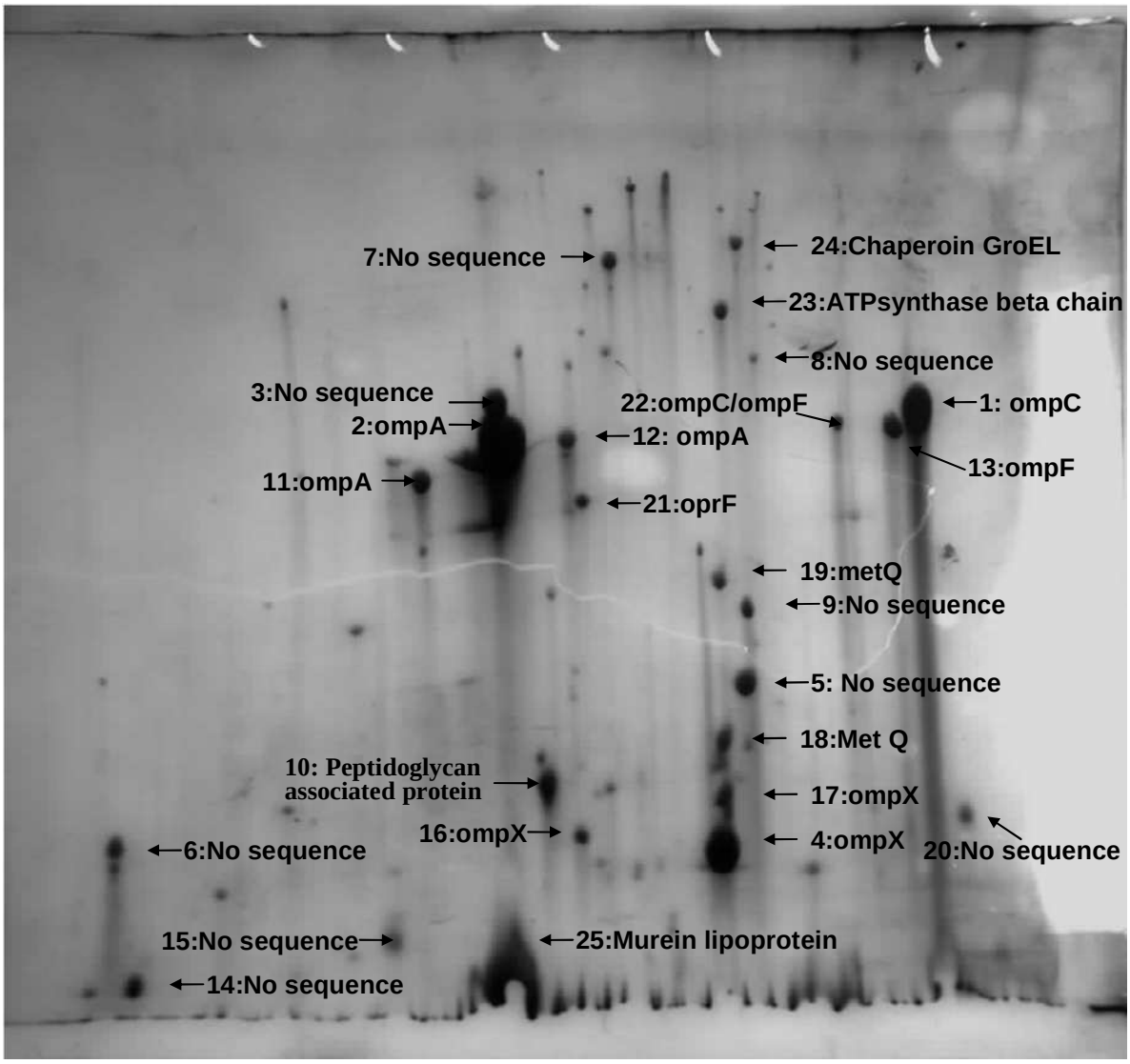


Figure 3-6: 2D–gel of membrane proteins from Iso10 identified by DeNovo sequencing and PMF.

pH7 **6** **5** **pH4**

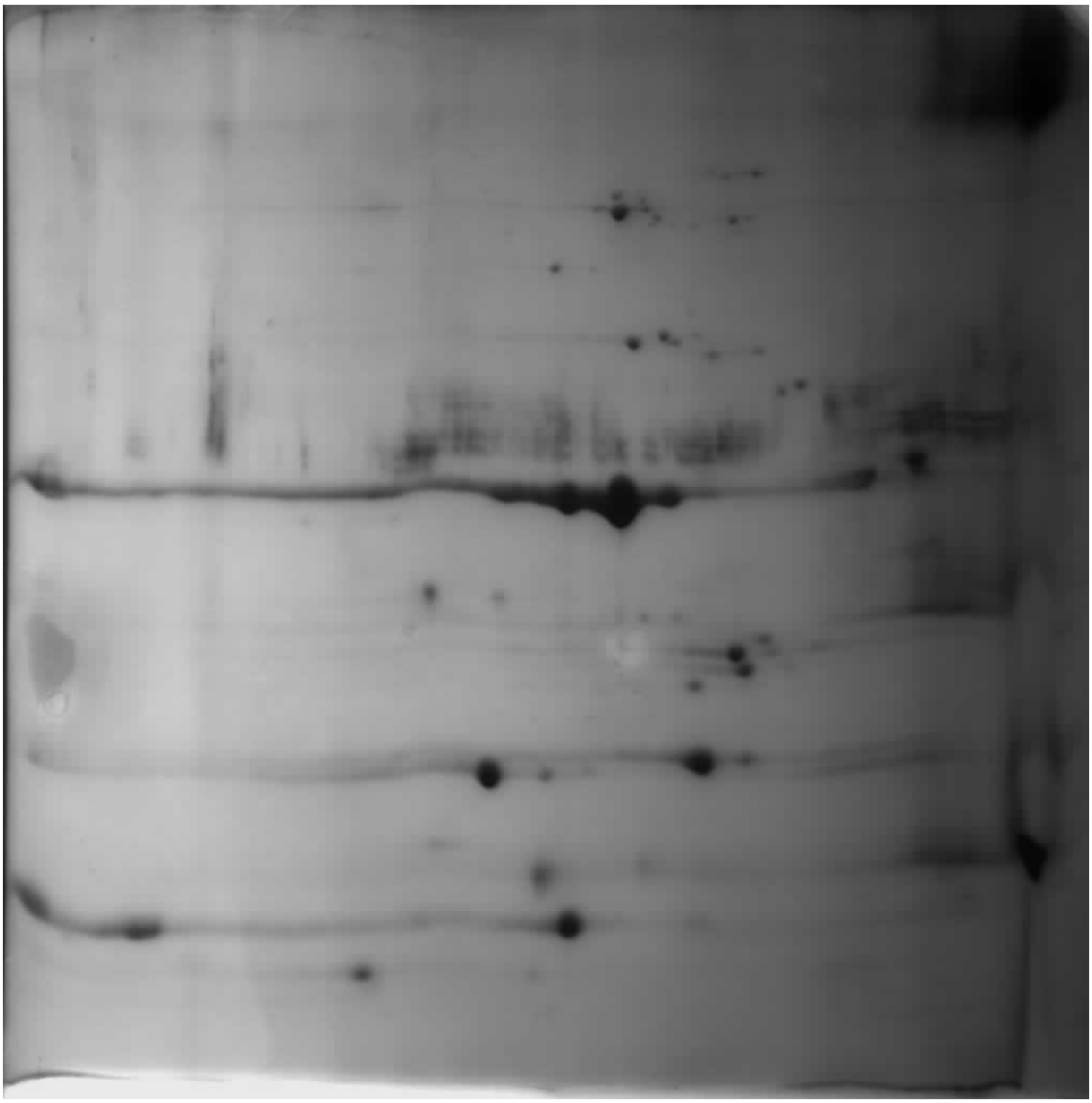


Figure 3-7 : 2D gel with membrane proteins from the *Shigella boydii* type strain.

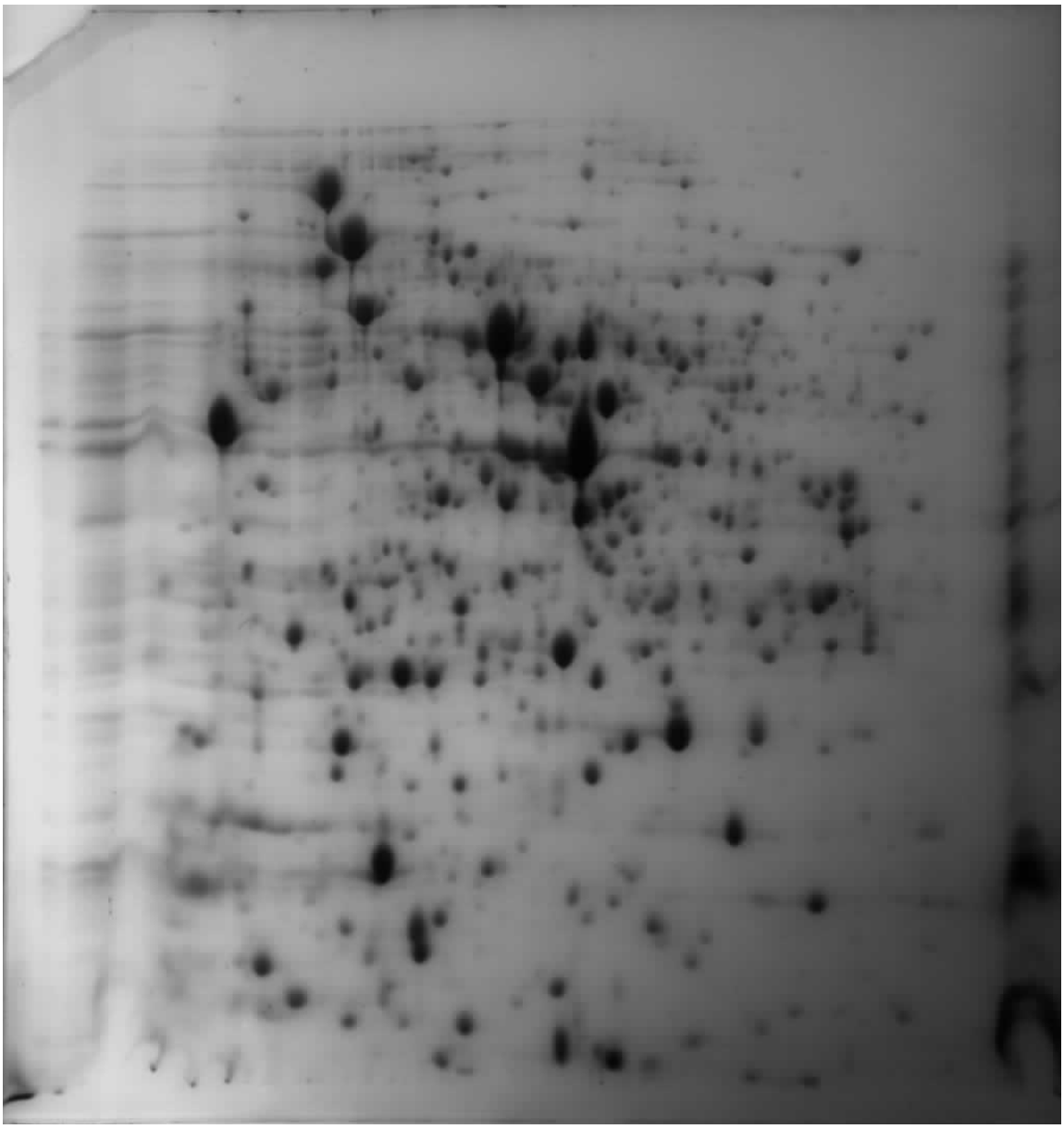


Figure 3-8: 2D gel with soluble proteins from the environmental isolate Iso10.

Table 3-8 Outer membrane proteins identified by DeNovo sequencing and search in blastp (NCBI)

Number	CAF – Fragment	Sequence	NCBI Acc. Number	E–value	Protein common name	Gene Symbol	Function |	Species
1	1333,46	NVGDAYGFANK	XP001757869	18	hypothetical protein			*Physcomitrella patens subsp*
	2383,07	NTDFFGLVDGLNFAAQYQGK	ZP02779974	2E–07	Outer membrane protein C	OmpC	Porin, Cell envelope	*Salmonella enterica*
2	970,31	I/LGGFVWR	AAK68970	274	Outer membrane protein A	OmpA	Porin, Cell envelope	*Pectobacterium chrysanthemi*
	1261,47	SDVI/LFNFNK	ZP02812396	3.4				*Enterobacter aerogenes*
	2256,81	I/LGWSQFHDTGFYGDGYDK	YP001477986	3E–05				*Serratia proteamaculans*
4	1234,37	FKSTDYPAK	XP_507976	315	ATP-binding cassette		Acid/base induced putative adhesin	*Pan troglodytes (chimpanzee)*
	1552,55	GQYYGI/LTAGPAYR	AAP22174	0.25	Outer membrane proteinX	OmpX		*Enterobacter aerogenes*
	1642,64	NVDVGTWI/LAGVGR	AAP22174	0.007		OmpX		*Enterobacter aerogenes*
10	1248,5	VTVEGHADER	YP2636366	0.57	Peptidoglycan associated outer membrane lipoprot.			*S. enterica*
	1471,75	GTPEYNI/LSI/LGER	YP1175971	0.13				*Enterobacter sp*
	1601,8	GVSADQI/LSI/LVSYAGK	ZP03075787	0.011				*S. enterica*
12	1189,42	DNTWYAGGK	YP858623	4.2	Outer membrane protein A	OmpA	Porin, Cell envelope	*Aeromonas hydrophila*
	1261,52	SDVI/LFNFNK	ACN65406	3.4				*Edwardsiella tarda*
13	1223,5	I/LEQWATGI/LK	ZP02903655	28	Outer membrane protein F	OmpF	Porin, Cell envelope	*Escherichia albertii*
	1327,4	FGDAGSFDYGR		1.7	Omp36/OmpC/OmpF*			
	1652,62	NSTYI/LSDSAGFANK	ZP02084037	36	Hypothetical protein			*Clostridium bolteae*
17	Unknown	GQYYGI/LTAGPAYR	YP001453862	0.005	Outer membrane protein X	OmpX	Putative adhesin	*Citrobacter koseri*
18	1042	DGI/LFVEGK	YP944818	82 (8/8)	DL-methionine transporter substrate binding subunit	MetQ	D–methionine binding lipoprotein	*Psychromonas ingrahamii*
19	2040,6	FVEAYQSDEVYEAANK	ZP02903098	5E–06		MetQ		*S. enterica*
21	1539,6	YYFTNNI/LYAR	AAR01584	0.066	Outer membrane protein	OprF	Porin, Cell	*Pseudomonas*

				(10/10)	OprF		envelope	*marginalis*
	1561,6	TNNI/LYAR	AAR01584	232 (7/7)	OprF			*Pseudomonas Marginalis*
	2335,0	…ADNATEAGR		21	OprF/OmpA-family protein**			*Pseudomonas sp /Burkholderia cenocepacia*
22	1464,4	YWVGSFDYGR		57	OmpC/OmpF***		Porin,Cell envelope	
23	1581,8	DVI/LI/LFVDNI/LYR	YP2706036	0.29	ATP synthase beta chain, mitochondrial, putative, F1 sector	AtpD	Energy metabolism	*Stenotrophomonas sp*
	1665,8	YTI/LAGTEVSAI/LI/LGR	YP002007073	8E–04				*Cupriavidus taiwanensis*
	2121,0	N(184)I/LEHSGYSVFAG VGER****	YP001174655	2E–04				*Pseudomonas stutzeri*
	22.04,1	NI/LASI/LGI/LYPAVDP I/LDSTSR	YP001785168	1E–06				*Haemophilus somnus*
24	1783,0	ACEEGVVAGGGVAI/LI/LR	YP002003184	0.002	chaperonin hsp60GroEL	GroEL/MopA	Protein folding	*Neisseria gonorrhoeae*
25	1446,6	I/LDNQAHSYQR	YP001335792	32	Outer membrane lipoprotein precursor (Murein-lipoprotein)	Lpp	Cell envelope	Klebsiella pneumoniae

***100% (11/11) match with three different proteins: OmpF, OmpC and Omp36porin.**
**** 100% (9/9) match with both Opr F and OmpA- family protein.**
***** 100% (8/8) match with both OmpF and OmpC**
******The number 184 (spot 23) most likely represent the mass of two amino acids that are not possible to differentiate from each other.**
From a dipeptide mass table: - 184 Da may represent P and S or I/L and

Table 3-9 Outer membrane proteins identified with PMF

Spot nb	Exp pI/Mr	Obs. pI	PBMS score*	Acc Nb	Description
11	5.99/37292	5.75	154	gi|15800816	OmpA, *E. coli* O157:H7
16	5.04/16350	5.5	97	gi|6435772	OmpX *Enterobacter cloacae*

*** Spot 11: Protein scores greater than 77 are significant (p<0.05).**
Spot 16: Protein scores greater than 78 are significant (p<0.05)

Unfortunately no 2D gel with membrane proteins resuspended in membrane solubilising solution 2 from *S. boydii* was achieved and a direct comparison with Iso10 was therefore not possible. The quality of the only 2D gel with membrane proteins from *S. boydii* (Figure 3-7) was quite poor and a number of proteins might not be visible due to streaking.
Some of the membrane proteins like OmpA and OmpX resolved into multiple charged isoforms. This has earlier been reported from other 2DE studies which use the carbonate extraction method (Molloy et al. 2000; Ying et al. 2005).
The sequences showed homology with membrane proteins from a range of bacteria from species in the *Escherichia* genus to far more distantly related bacteria like *Cupriavidus taiwanensis.* This implies that the membrane proteins are highly conserved even among distantly related bacteria.

Running of 2D gels with soluble proteins isolated without precipitation was not successful and only showed streaking on the polyacrylamide gel.
By precipitating the proteins with TCA in acetone and using 2D clean up kit to reduce streaking a well focused gel was produced with only minor streaking at the acidic end of the gel. One 2D gel with soluble proteins from Iso10 was produced (Figure 3-8) but the procedure failed to give representative results when precipitation of proteins from *S. boydii* was performed. Since new arrangements with Haukeland Hospital would be needed for the cultivation of more *S. boydii* cells and there was only a short time left of the study it was decided to focus on the membrane proteins of Iso10 and not repeat the experiments with *S. boydii* proteins.
Since the gel showing protein expression of soluble proteins from Iso10 could not be compared with a directly comparative gel of *S. boydii* it was also decided to primarily focus on identification of membrane proteins from Iso10.

3.5 Western blot of membrane proteins

2D gels with membrane proteins from Iso10 and *S. boydii* were blotted onto cellulose membranes and brought to University of Dhaka in Bangladesh for hybridisation with *S. boydii* specific antisera. However, after hybridisation, no visible spots could be detected on the membranes and time limitation did not allow me to repeat this experiment. Antiserum was brought back to Norway and used for Western blot analysis of membrane proteins from Iso10 separated by SDS–PAGE (Figure 3–9) and blotted to new cellulose membranes.
The hybridisation of these membranes was successful and several bands were revealed, demonstrating the presence of cross-reacting immunogenic proteins (Figure 3–10).
The lower part of the blot was covered with a smear in an area where no proteins were visible on the SDS–PAGE stained with Commassie Blue. This can be caused by traces of LPS in the sample and since coomassie does not stain LPS it would not be visible on the gel. One strong band and 7 weak bands representing immunogenic proteins (marked with arrows) could be distinguished. Some, but not all of the weak bands could be correlated with bands on the SDS–PAGE as stained with Comassie blue. Two of the membrane proteins stood out as very strong bands on the SDS–page (35–37kD), but only a single band corresponding to one of these abundent proteins could be detected in the Western blot, indicating a strong difference in their immunogenic properties. Although the two strong bands have nearly the same molecular mass, they can easily be cut out of the gel separately for further analysis.

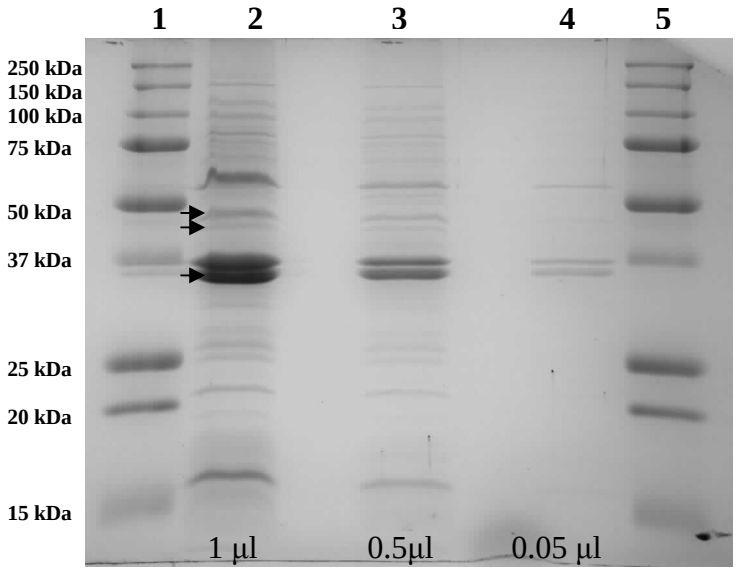


Figure 3-9: Dilutions of membrane proteins from Iso10 stained with coomassie. Protein bands correlating with bands on the western blot are marked with an arrow.
Lane 2, 3 and 4 contain 1µl, 0.5µl and 0.05µl of the membrane protein extract respectively. Lane 1 and 5 contain the All blue standard (Biorad 161–0373).

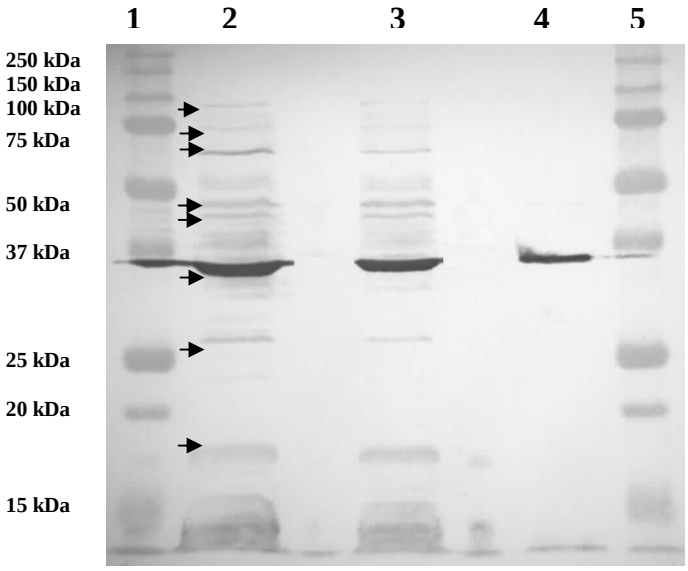


Figure 3-10: Western blot of membrane proteins from Iso10 probed with *S. boydii* 15 specific antisera (diluted 1:1000). Eight bands of immunogenic proteins are marked with an arrow. Lane 2, 3 and 4 represent western blot of 1µl, 0.5µl and 0.05µl of the membrane protein extract respectively. Lane 1 and 5 contain the All blue standard (Biorad 161–0373).

3.6 Identification of immunogenic proteins

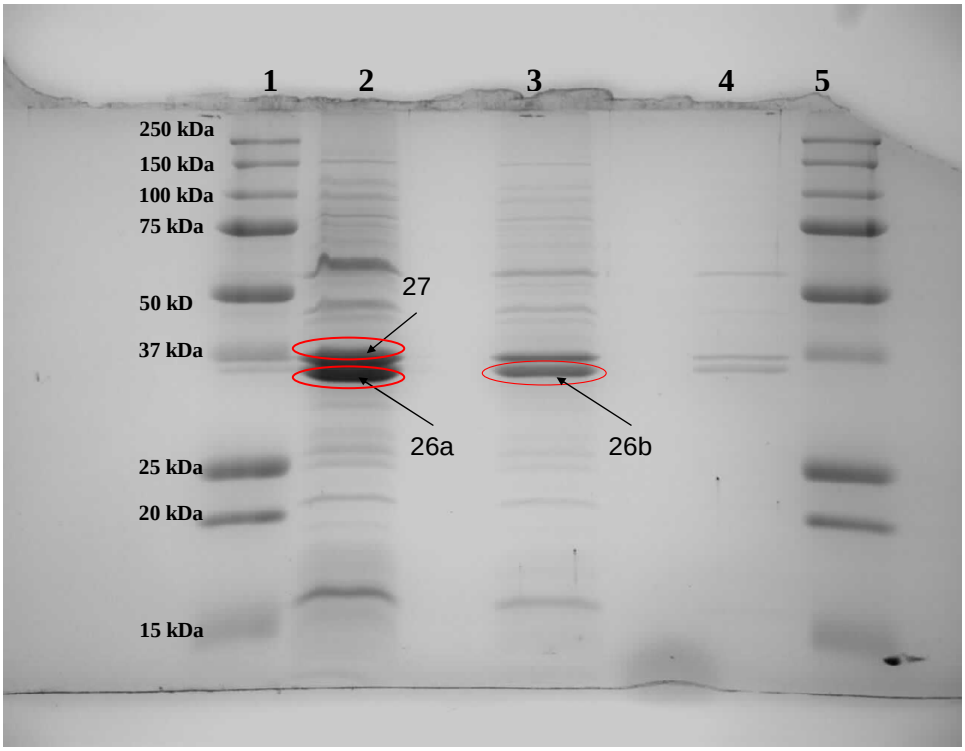


Figure 3-11: SDS–page with membrane proteins from Iso10. Excised bands are marked with red circles and spot number. Lane 2, 3 and 4 contain 1µl, 0.5µl and 0.05µl of the membrane protein extract respectively. Lane 1 and 5 contain the All blue standard (Biorad 161–0373).

Table 3-10: Identified membrane proteins.

Spot number	CAF–fragment	Sequence	NCBI acc. no	E–value	Protein
26a	1478,7	FWYGSI/LDYGR	*	42	
	1495,8	VVI/LGYTDR	ZP_02805910	57	OmpA
	1870,0	I/LPITDDI/LDI/LYTR	ZP_02805910	0,020	OmpA
26b	1478,3	FWYGSI/LDYGR	*	42	
	1870,0	I/LPITDDI/LDI/LYTR	ZP_02805910	0,020	OmpA
27	1478,3	FADYGSI/LDYGR	YP_069796	0,049	OmpC
	1500,3	CADYGSI/LDYGR	YP_001871782	0,62	
	1549,4	I/LDDNNFTR	ABY81620	19	

*** Give 100% match (8/8) with the following proteins: NmpC precursor, ompC, ompF and other unspecified gram negative outer membrane porins**

The two strong bands marked as spot 26 (a and b) and spot 27 in Figure 3-11 were excised and identified with deNovo sequencing (Table 3-10). Spot 26, which gave a strong, specific cross-reactivity with *Shigella*-specific antisera was identified as outer membrane protein A. The second spot, with a higher molecular weight and no antigenic cross-reaction, was identified as outer membrane protein C. The Iso10 OmpA protein is thus identified as the main cross-reacting protein.

3.7 Comparative studies of OmpA from E. coliK12, E. fergusonii, S. boydii and Iso10

In order to assess the specificity of the OmpA cross-reaction, a Western blot comparison of 4 different strains was carried out. Membrane proteins from *E. coli* K12, *E. fergusonii* ATCC 35469, *S. boydii* ATCC12034 and Iso10 were isolated, separated by SDS-PAGE and used for Western blotting with *S. boydii*15 type-specific antisera.

The SDS–PAGE from E. *fergusonii* and *E. coli* K12 showed two bands of membrane proteins with almost the same molecular weight as the identified OmpA and OmpC in the Iso10 sample. The sample with membrane proteins from S. *boydii* ATCC only showed one band in the corresponding area (Figure 3–12).

The Western blot (Figure 3–13) showed one band cross- reacting with the type specific antisera in all the four samples. Surprisingly, the crossreaction was weaker for the *S .boydii* protein than for the other three samples. The cross- reacting proteins in Iso10, *E.fergusonii* and *S. boydii* samples had same molecular size (approximately 36kDa).

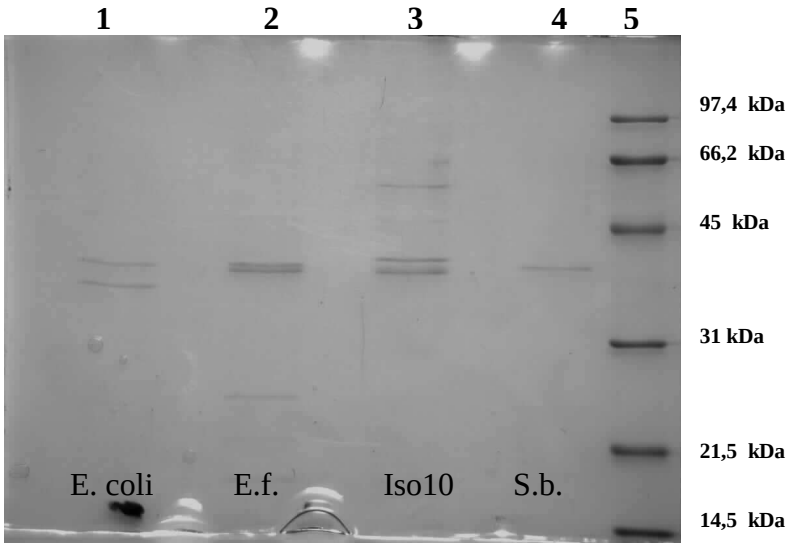


Figure 3-12: SDS–page with membrane proteins stained with coomassie. Lane 1, 2, 3 and 4 contain membrane extract from *E. coli* K12, *E. fergusonii*, Iso10 and *S. boydii* respectively. Lane 5 contains a High range standard (BIORAD 161–0303).

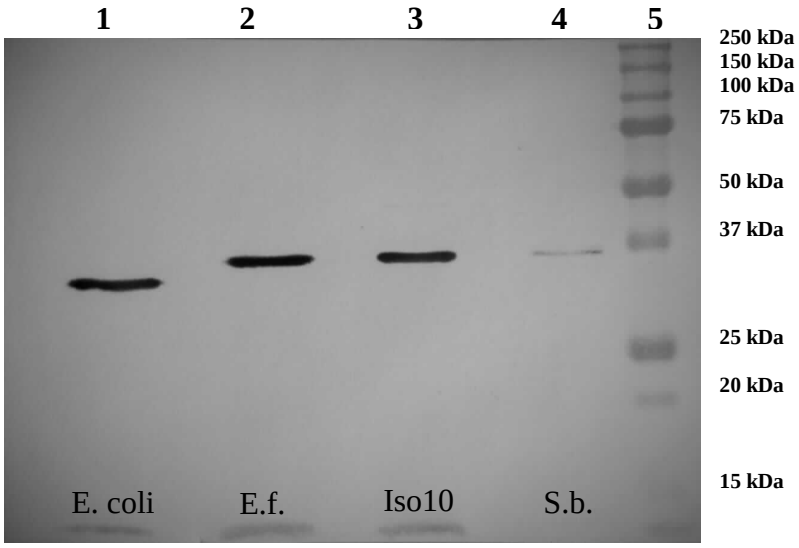


Figure 3-13: Western blot of membrane proteins. Lane 1, 2, 3 and 4 contain membrane proteins from *E. coli* K12, *E. fergusonii*, Iso10 and *S. boydii* respectively. Lane 5 contains the All blue standard (Biorad 161–0373).

3.8 Western blot of LPS

In order to investigate the possible role of LPS in the serological cross- reaction between Iso10 and *S. boydii* a comparative Western blot analysis was carried out. LPS was extracted from *S. boydii* 15, *E. fergusonii* ATCC 35469 and Iso10 and separated by SDS-PAGE. Silver-staining confirmed that the LPS extraction was successful (Figure 3–14). Western blotting using the *S. boydii* 15 antiserum clearly showed that Iso10 and *S. boydii* LPS cross-react strongly, confirming previous results using whole cells extracts (Rahman et al. 2007). However, no cross-reaction could be detected with of LPS from *E. fergusonii* (Figure 3–15). This clearly shows that Iso10 differs strongly from the *E. fergusonii* type strain with respect to O-antigenic properties, and indicates that LPS is the main reason for the strong cross-reaction with *S. boydii.*

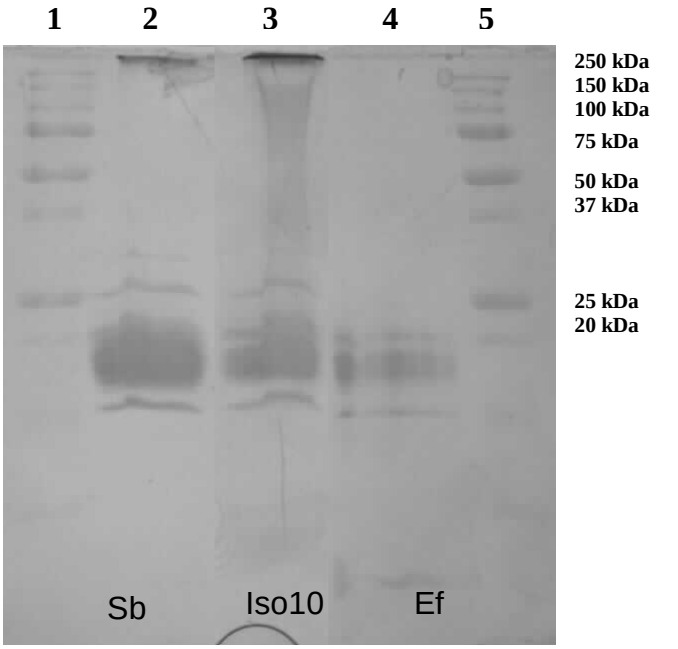


Figure 3-14: Silver stained SDS-PAGE gel with LPS. Lane 2, 3 and 4 contain LPS isolated from *S. boydii*, Iso10 and *E. fergusonii* respectively. Lane 1 and 5 contain the All blue standard (Biorad 161–0373)

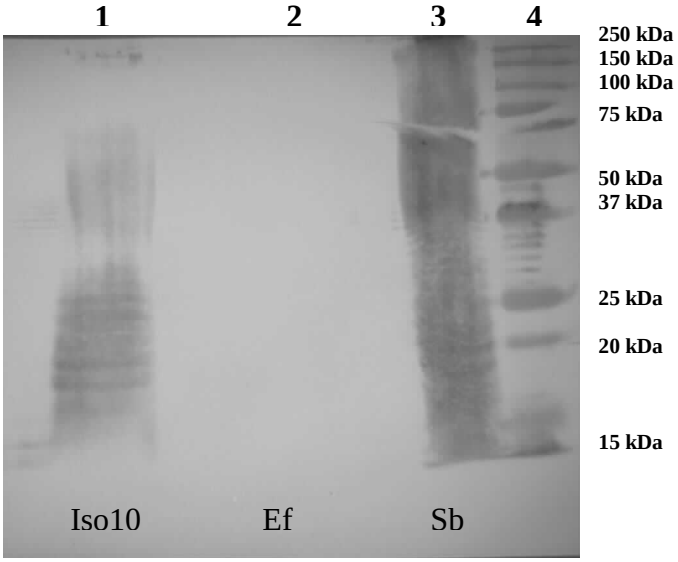


Figure 3-15: Western blot of LPS. Lane 1, 2 and 3 represent western blot of LPS from Iso10, *E. fergusonii* and *S. boydii* respectively blotted against *S. boydii* 15 type-specific antisera. Lane 4 contains the All blue standard (Biorad 161–0373).

4 Discussion

Diarrhoeal diseases caused by *E. coli* and *Shigella* spp. are a major health problem today, especially in developing countries. Extensive studies of the virulence and epidemiology of *Shigella* has been carried out, but even though contaminated water is thought to be important for the dissemination of *Shigella*, its survival and distribution in the aquatic environment is still not understood. Only one report has earlier described the isolation of *Shigella* from environmental water (Faruque et al. 2002).

This project has focused on a bacterial isolate (strain Iso10) from a lake outside Dhaka in Bangladesh. An earlier study has revealed that Iso10 efficiently cross-react with *Shigella boydii* 15 type-specific antisera but does not possess the genes associated with *Shigella* virulence, and thus represents a potential live *Shigella* vaccine (Rahman et al. 2007).

This work was divided into three main parts; (1) A MLST study to identify the genetic relationship between Iso10, *Shigella* and *E. coli*, (2) a proteomic study to compare the protein characteristics of Iso10 with *S. boydii* 15 and (3) an immunogenic analysis of the cross-reaction between membrane proteins and LPS from Iso10 with *S. boydii* 15.

The MLST analysis of seven housekeeping genes revealed a close genetic relationship between Iso10 and *E. fergusonii.* The close relationship with *E. fergusonii* was supported by biochemical characterization with API20E.

Interestingly, Western blot of LPS from Iso10, *S. boydii* and *E. fergusonii* showed that only Iso10 and *S. boydii* reacted with type specific *Shigella* antisera, indicating an O-antigenic relationship between the two strains.

A 2D gel-electrophoretic analysis of membrane proteins from Iso10 was performed and ten different proteins were identified with deNovo sequencing. One of these proteins was identified as immunogenic and cross-reactive.

4.1 Discussion of methods; Multi Locus Sequence Analysis

The ability to accurately discriminate different strains of infectious bacteria is crucial for epidemiological surveillance and phylogenetic studies in microbiology. Several typing techniques have been used in order to distinguish strains of bacteria within the same species, including Pulsed Field Gel Electrophoresis (PFGE), Multi Locus Enzyme Electrophoresis (MLEE), serotyping, ribotyping, Random Amplification of Polymorphic DNA (RAPD),

Amplified Fragment Length Polymophism- PCR (AFLP-PCR) and Multi Locus Sequence Typing (MLST)(Maiden et al. 1998).

PFGE is based on rare-cutting restriction enzymes and separation of the fragments by electrophoresis to provide a comparative genetic fingerprint. RAPD is a type of PCR- reaction where unspecific primers are used to amplify genomic DNA and make a comparative pattern. In ribotyping DNA is cut with restriction enzymes, separated in a gel, transferred to a membrane and probed with a rRNA gene probe which gives a unique pattern of bands. MLEE, which is the predecessor of MLST, distinguishes the relative electrophoretic mobility of the gene products from multiple genes (Cooper and Feil 2004).

The increased automation and reduced costs of DNA sequencing has revolutionized molecular typing techniques and taxonomic studies. The development of online databases has made specific DNA sequences from microorganisms from all over the world available. This was essential for the choice of methods since a sufficient amount of data is vital to determine correct taxonomic relationship between Iso10 and representatives of the *Escherichia* and *Shigella* genera. Most of the techniques listed earlier are image based DNA – fingerprinting which use comparative DNA fragment patterns made by PCR techniques or restriction enzymes. Even though much work has been done to standardize the protocols and interpretations, comparisons between laboratories are still difficult. AFLP-PCR is in this case unique because it combines both of these elements by digesting genomic DNA by restriction enzymes and use selective PCR- amplification to yield a specific electrophoretic fingerprint. The technique is in many studies described as very effective with high discriminatory power. However, for long term epidemical surveillance, exchange of data between laboratories is most important and even though AFLP is more reproducible than many of the techniques based on amplification or restriction, it is less reproducible than sequence-based typing like MLST (Melles et al. 2007). In my case, MLST makes the laboratory work considerably more effective since sequences from *Escherichia* members for the chosen genes already were available online for comparisons with my isolate.

Since the MLST method is able to identify differences down to single bases it provides high discrimination power, and has proven to have higher resolution level than its predecessor, MLEE (Cooper and Feil 2004). Nevertheless, fingerprinting based on restriction cutting like PFGE is able to identify more large scale genomic rearrangements in strains with mobile elements like for example insertion sequences and genomic islands.

Another important issue, especially in immunology, is “serotype switching”. This is a mechanism were horizontal gene transfer of the capsular genes takes place and has earlier

been reported in *Streptococcus pneumoniae* (Coffey et al. 1998) and *meningococcal* (Arreaza et al. 2003) strains. Numberous of reports has also observed horizontal gene transfer of the genes encoding the O- antigen in gram- negative bacteria (Fegan et al. 2006). These strains may be classified as not closely related by MLST, but share identical serotype. On the other hand, studies which are only based on serotyping may identify bacteria incorrectly. It is crucial to identify serotype switching especially in vaccine studies and MLST should therefore always be combined with serotyping (Cooper and Feil 2004).

Based on earlier studies, Iso10 was identified as belonging to the *Shigella*/*E. coli* group. The best match from database search with the V3 16S rRNA gene region sequence was *E. coli*, but serologically it was identified as *S. boydii* 15 (Rahman et al. 2007).
Based on these observations it was decided to use the Achtman MLST database for *E. coli* (including *Shigella*) to investigate the genetic relationship between Iso10 and representatives of different pathogenic groups of *E. coli* and *Shigella.*
The MLST database for *E. coli* uses seven housekeeping genes distributed around the chromosome for phylogenetic identification of *E. coli* strains. Housekeeping genes are present in all cells because they provide basic functions, often associated with central metabolism. They evolve at moderate rate providing sufficient information for intra-species comparisons but will not lead to too diverse sequences which often is the case with hyper-variable loci, like certain antigenic proteins (Cooper and Feil 2004).

The genes were PCR amplified and directly sequenced. The primers described in the Achtman database were used to amplify all the genes except *mdh* which could not be amplified by the standard primer set. A PCR product was finally obtained after many attempts by using primers designed by Pupo et al (1997). This primer set is similar to the Achtman primers but shorter (18-20 nucleotides), which allows a lower annealing temperature as compared with the other primer set, which consisted of as much as 35 nucleotides. By studying the search results in Table 3.2 it also becomes clear that the *mdh* gene in Iso10 and *E. fergusonii* (100% identity) only shares 87% identity with *E. coli.* This may also explain why the PCR reactions were unsuccessful since both of the primer sets were designed for amplification of genes of *E. coli.* A shorter primer will be less specific allowing amplification of the gene from more distantly related strains.

The seven different sequences were put together to one long concatenated sequence and aligned with corresponding concatemers of representatives from the different pathogenic groups of *Shigella* and *E. coli*.

4.2 Discussion of methods; 2D Electrophoresis

2-D gel electrophoresis is today the only technique that can be used for parallel, quantitative expression analyses of large sets of complex proteins mixtures. By combining isoelectric focusing (IEF) with SDS-PAGE the protein mix is separated according to isolectric point (pI), molecular mass (M_r) and solubility (Gorg et al. 2004). This gives the opportunity to easily get a visual overview of the proteins of interest (soluble or hydrophobic) and an increased separation compared to the classic, one dimensional SDS-PAGE electrophoresis.
The original protocol for 2D electrophoresis (O'Farrell 1975) was improved by the introduction of immobilised pH gradients (IPG) (Gorg et al. 1988) which replaced the tube gels. This method made the isoelectric focusing more precise and made it possible to separate proteins in much wider portions of the pH scale from very acidic to very basic. It also improved the reproducibility, resolution and sample loading capacity which had been some of the major limitation of IEF. 2-DE can now resolve more than 5000 proteins simultaneously and detect protein amounts of less than 1ng per spot (Gorg et al. 2004).
The later years more programmes for 2D- image analysis have become available and online databases to analyse data from PMF with MALDI-TOF MS or deNovo sequencing with MALDI TOF-TOF MS have been extended and improved.
Even though the overall quality of 2D gels have been exceedingly improved since the first introduction by O' Farrell the technique still has some limitations, especially when it comes to membrane proteins. It was early noticed that the resolution of some classes of proteins, especially hydrophobic proteins, was poor or almost absent from the 2D gels (Wilkins et al. 1998). A modified version of the carbonate extraction protocol from Fujiki et al (1982) has been successfully used to extract outer membrane proteins (OMP) from *E. coli* (Molloy et al. 2000) and *S. flexneri* (Ying et al. 2005). The two different OMP studies used different solubilisation solutions but succeeded in identifying 55-58 potential OMP's, constituting approximately 80% of the predicted integral membrane proteins. On the basis of these studies it was decided to use the carbonate extraction method for isolation of membrane proteins. Three different membrane solubilisation solutions were used, all including the chaotropes urea (7M) and thiourea (2M) but with different combinations of detergens and reducing

agents. The quality of the gel and solubilisation of the membrane proteins varied significantly depending on the solubilisation solution chosen.

The first solution, which is routinely used for solubilisation of soluble proteins, with detergents CHAPS and Trixton X-100, and DTT as reducing agent, gave very poor results with only 6-8 protein spots (picture not shown). The second solution with TBP as reducing agent, detergent ASB-14 with Tris-base gave better resolution but horizontal streaking decreased the quality of the gel (Figure 3-7). The third solution with DTT as reducing agent and ASB-14 as detergent gave the best focused gel with 52 visual spots (Figure 3-6). Both DTT and TBP are commonly used for solubilisation of membrane proteins. TBP has in some cases lead to improved solubility and identification of proteins that were previously unresolved when DTT was used (Molloy 2000). The reason for horizontal streaking in this case remains unclear but it can be caused by non-protein impurities in the sample.

Two different methods were used to extract soluble proteins. Extraction by TCA/acetone precipitation and extraction with lysis buffer without precipitation. Extraction by TCA/acetone precipitation is the most commonly used method and is very efficient for removal of interfering compounds and minimisation of protein degradation. However, protein loss due to incomplete precipitation is quite common and the final gel may not reveal all the proteins present in a whole cell lysate. Considering this, an alternatively extraction method without precipitation was also used. This method was not successful and the gel only showed hoisontal streaking with no spots. So until an alternative method is developed the TCA/acetone precipitation method seems to be the best technique to provide extracts for 2DE gel analysis of whole cell lysates.

Recently, newer methods for comparative proteomics have emerged, including multi-dimensional chromatography/liquid chromatography – tandem mass spectrometry (LC/LC–MS/MS) with isotope-coded affinity tags (ICAT) for accurate quantification of the components of the proteins by MS (Gygi et al. 2002). This is a promising technique where the amount of data for the analysis is less restricted and the detection of certain groups of proteins is more efficient than 2DE, like for example very basic/acidic proteins and membrane proteins. It also detects low abundance proteins whereas 2DE only detects medium-and high abundance proteins.

However, protein changes which often are detected on 2-D gels can be impossible to detect by using the ICAT methods. The huge growth of the 2DE field combined with the excellent resolution and the ability to store the proteins in the gel still makes 2DE one of the most common and important separation techniques in proteomics (Lopez 2007).

Gel to gel variance is a common problem in 2DE when two gels are compared by computer image analysis programs. To overcome these problems a method called Difference Gel Electrophoresis (DIGE) was developed. The principle for this method is that two samples (e.g. proteins from a strain cultivated at two different temperatures) are labelled with different fluorescent dyes (CyDyes), which excitate at different wavelengths, are mixed and applied at the same 2DE gel (Gorg et al. 2004). The spots from each strain will be shown in two different colours and the overlapping spots will be shown in an other colour. Since comparative gels were not obtained in my experiments DIGE was never an option as a comparative technique.

Peptide Mass Fingerprinting (PMF) by using MALDI TOF MS is the most common method for protein identification. It measures the masses of tryptic peptides by MS which gives a unique PMF and matches this with the theoretical peptide masses in a database. It is a very effective tool for protein identification from species where the genome is small and completely sequenced, but less reliable for species which are less studied. It also does not give good results if one spot contains more than one protein. Since the identification of Iso10 still was unclear when the experiments started I chose deNovo sequencing as an alternative method for protein identification. This method does not rely on a complete genome sequence of the species and give a specific sequence which can be compared with the sequence information from other species. The major drawback of this technique is that it is time-consuming and requires expertise to analyse the MS/MS spectra.

Twenty-five spots were excised from the 2D-gel and sent for analysis by MALI TOF/TOF. Spots that did not give a protein sequence were also analysed by using the PMF technique. Fourteen spots were identified with deNovo sequencing and 2 spots with PMF, identifying 10 different proteins. The identification of the first ten spots (spot 1-10) failed and all the samples were lost, only a few of them (spot 1, 2, 4 and 9) could be analysed again with new samples. This partly explains why only 64% of the original samples were fully identified. The success rate of identification with a combined PMF and MALDI TOF/TOF MS analysis is normally 75 % to 95% (Suckau et al. 2003). By excluding the first unsuccessful attempt, 16 out of 20 (80%) samples were identified showing that our identification rate is in accordance with the expectations by this method.

A parallel 2D gel of the membrane proteins was electroblotted to a cellulose membrane and brought to Dhaka University for hybridisation with *Shigella*- specific antisera. Immunoreative proteins were only visualized on a cellulose membrane blotted from a small SDS-PAGE but not from the 2D gel. The reason for the unsuccessful hybridisation of the 2D gel is still

unclear and the procedure was too time-consuming to repeat. Two different electroblotting techniques were used for SDS-PAGE and 2D gel. Even though the all blue standard were visible on the membrane, the proteins might not have been successfully transferred from the 2D gel. The gels might also have become damaged during transport e.g. due to heat.

4.3 Discussion of results; Multi Locus Sequence Analysis

The 16S rRNA gene is distributed universally among bacteria and is routinely used to indentify bacteria and characterize complex microbial communities. However, the 16S rRNA gene evolves so slowly that discrimination of closely related strains, especially among the γ-proteobacteria, based on this gene alone may not always be possible.

The previous studies of Iso10 only included a partial sequencing of the 16S rRNA gene. In my work I sequenced the complete 16S rRNA gene (1429bp) and used it for database search. It showed high similarity to *E. fergusonii, E. coli* and *Shigella* spp. but could not distinguish between the three different species. From this we could only conclude that Iso10 is in the *Escherichia* (including *Shigella*) genus, but the 16S rRNA gene did not give sufficient information to identify it at the species level.

To be able to place Iso10 more specifically in the *Escherichia/Shigella* group a multi locus sequence analysis was carried out based on seven genes (*purA, adk, icd, fumC, recA, mdh* and *gyrB)* evenly distributed around the chromosome.

The sequences were first used for search in the online sequence databases. Search by using complete sequences gave identity values from 85% to 99% to members of *Escherichia*. These database searches were done before the whole genome sequencing project of *E. fergusonii* was completed and published. Later search with short sequences gave 99% – 100% identity match to *E. fergusonii* for all the genes (Table 3-2).

Wirth et al. (2006) reported in the study of 420 *E. coli* isolates that within each of the seven chosen gene fragment 8–20% of the nucleotides were polymorphic (Figure 4–1) (Wirth et al. 2006). The search results from Iso10 follow the same pattern of polymorphism. These data showed that sequences from the genes *recA* and *purA* are most conserved while sequences from *fumC* are most polymorphic. R*ecA* and *purA* sequences from Iso10 share 98– 99% sequence identity with *E. coli* while the *fumC* sequence only shares 85% sequence identity to the best match, excluding *E. fergusonii*. However, the *mdh* sequence from Iso10, which has a 100% identity match with *E. fergusonii*, only shares 87% sequence identity with the

representatives of *E. coli,* while the data from Wirth et al. show that the level of nucleotide polymorphism in *mdh* from *E. coli* strains are almost at the same level as *recA* and *purA.*

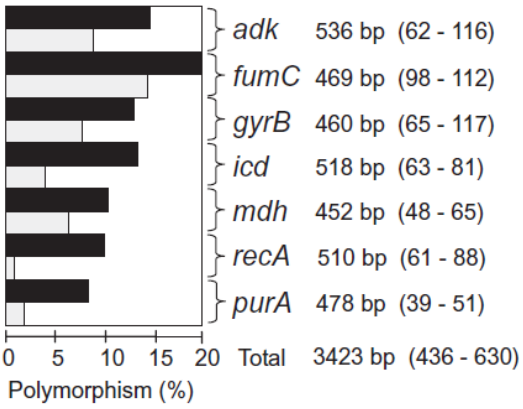


Figure 4-1: Levels of polymorphism in each of the genes presented in a histogram. Black bars represent nucleotide polymorphism and gray bars represent amino–acid polymorphism. Each symbol is followed by the length of the sequenced gene fragment (informative sites- polymorphic sites) (Wirth et al., 2000).

The sequences from Iso10 were also used in Multi Locus Sequence Typing by using the MLST database for *E. coli*. The analysis showed that Iso10 shared high sequence similarity with four of the genes, *recA* (99.4%), *purA* (99.8%), *icd* (99.0%) and *gyrB* (100%), from two sequence types (ST770 and ST485) (Table 3-4). However, no good allele matches were found for the *adk* (93.4%), *fumC* (85.4%) and *mdh* (87.5%) sequences, making the two closest database matches insignificant. Since Iso10 or any close relative obviously is not included in this database, which currently represents the biggest collection of alleles from *E. coli, Shigella* and even *E. albertii,* Iso10 is most likely not a member of the *E. coli*/*Shigella* group.
To clarify and make a visually presentation of the phylogenetic relationship of Iso10 and *Shigella, E. coli, E. fergusonii* and *E. albertii* a phylogenetic tree using the concatemeric sequences was constructed which included all the pathogenic and ancestry groups of *E. coli* (Figure 3.2). In the phylogenetic tree Iso10 and *E. fergusonii* was clearly placed together, outside the main cluster of *E. coli*/*Shigella.* The four sequence types (ST 485, 770, 662, 747), which were considered most closely related by the Achtman database, were placed in separate cluster outside the main cluster of *E. coli/Shigella.*
The tree in this work was not constructed to study the phylogenetic relationship of the pathogenic forms of *E. coli* and *Shigella.* This is already well covered by the scientific work of Pupo and his colleagues (Pupo et al. 1997; Pupo et al. 2000), and it would require a lot more data. However, some trends from these earlier studies can be recognised in the tree constructed here. The *Shigella* isolates fall into one of the main clusters with six other

representatives of *E. coli* falling between the four *Shigella* isolates. This supports the theory that *Shigella* is a group of closely related pathogenic *E. coli* strains. The data from Pupo et al. in 2000 showed three major clusters of *Shigella* and five forms not closely related to any of the other, suggesting that *Shigella* has arisen eight times and that the original taxonomy should be revised (Appendix H). They also found that the pathogenic groups of *E. coli* did not cluster neither with each other (except for *E.coli* O157:H7) nor with *Shigella* (Pupo et al. 1997). *S. dysenteriae* 1 and *S. sonnei* were two of the strains which in this analysis fell outside the main clusters of *Shigella*. *S. flexneri* strains and *S. boydii* 4, which were included in my tree, fell into two different clusters of *Shigella* (3 and 1, respectively) in the combined tree created by Pupo et al (Appendix H). This may explain why the *Shigella* representatives are not situated closer to each other in my tree. EIEC, which is most similar to *Shigella* in their pathogenesis, is not placed inside the *Shigella* cluster and *E. coli* O157:H7 was not placed close to the other EHEC strain. This also confirms the results from Pupo et al which showed that the pathogenic groups of *E. coli* did not cluster by using phylogenetic techniques. A correlation between pathogenic type and genetic structure may only be identified trough population-genetic tools, especially when homologous recombination has been a frequent event during evolution as suggested by Wirth et al (2000), who revealed that some virulence phenotypes were associated with certain ST complexes, including *Shigella* and EIEC.

4.4 Discussion of results; proteomics

By using carbonate extraction and solubilisation with strong zwitterionic detergents, membrane proteins were successfully extracted from Iso10 and separated according to isolectric point (pI) and molecular mass (M_r). A well focused 2D gel was obtained, visualizing 52 spots by coomassie staining. 25 spots were excised and sent for identification giving 16 positive results with 10 different proteins. Five of the proteins were outer membrane proteins (OmpA, OmpC, OmpF, OprF, OmpX), three were lipoproteins (peptidoglycan associated, Lpp and MetQ) and two were cytoplasmic proteins (GroEL/MopA and AtpD).

Outer membrane protein A, C, and F are well known hydrophilic channels which allow free diffusion of small molecules across the membrane. They are previously reported as the most abundant membrane proteins in both *E. coli* and *S. flexneri* (Ying et al. 2005; Weiner and Li 2008), with 10^5(OmpA), $2x10^4$ (OmpC) and 10^4 (OmpF) copies per cell in *E. coli*. The content of OmpC in *S. flexneri* is even higher that in *E. coli*. OmpX is also one of the

dominant conserved outer membrane channels but is also involved in the neutralizing of the host defence systems (Koebnik et al. 2000).

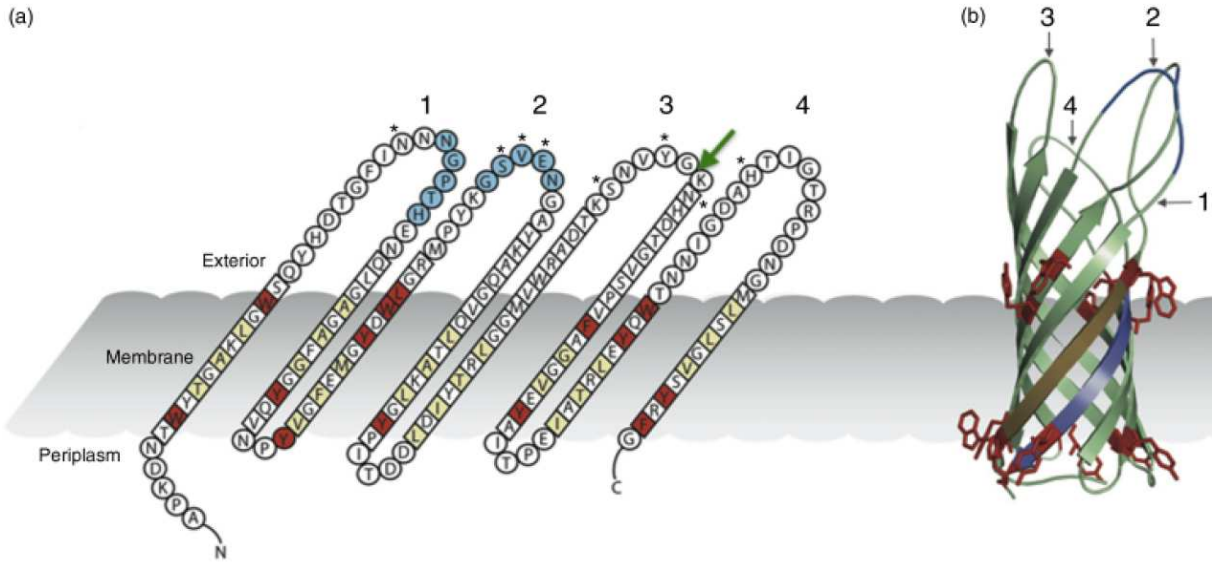


Figure 4-2 : The OmpA structure consisting of a predominantly β- stranded N- terminal domain with a globular C- terminal domain mainly made out of α-helixes. The porin is made out of eight antiparallels connected by four long loops. a) A topological model showing the major features of OmpA. b)Three dimentional structure of OmpA (Smith et al. 2007).

By using traditional SDS–PAGE and immunoblotting with *Shigella*- specific antisera, OmpA was identified as a strongly immunogenic protein (Figure 3–11). OmpC, with a slightly higher molecular weight, did not cross- react with the antisera. In earlier studies a strong immunogenic band (34-38kDa), thought to be a porin, was reported as a predominant antigen in all the four *Shigella* species (Roy et al. 1994a). My results confirm the data from Ying et al who identified OmpA as a strong immunogen and suggested it as a potential vaccine candidate protein (Ying et al. 2005).

OmpA (Figure 4-2) is in many cases compared with a Swiss army knife with its many functions and multifaceted properties and is highly conserved among the members of *Enterobacteriaceae.* It is critical for adherence and invasion of certain cell types and contributes to biofilm formation. It acts as an immunotarget and evasin and is also a receptor for several bacteriophages (Smith et al. 2007). In my experiments, OMPs from *E. coli, S. boydii, E. fergusonii* and Iso10 showed one strong band which reacted with *Shigella*– specific antigen in all the samples (Figure 3–13). This confirms the results from earlier studies implying that the OmpA gene is highly conserved among the *Escherichia* species. The Western blot showed identical pattern for OmpA from *E. fergusonii* and Iso10. The cross-reacting band from *E. coli* had a slightly lower molecular weight. Earlier studies of OmpA

from different species have shown that when protein samples with OmpA are boiled before they are run on an SDS-PAGE gel they undergo a conformational change and the molecular weight changes. The molecular weight after heating often varies in the range 33-36kD (Cole et al. 1982) and 35-38kD (Roy et al. 1994b), which may explain why the *E. coli* band in my experiment had a lower molecular weight. The cross-reacting band from *S. boydii* has the same molecular weight as Iso10 and *E. fergusonii* but is clearly weaker. This was not expected since the antiserum was *S. boydii* specific. However, when the protein concentration of the membrane extractions were measured it became clear that the concentration of the *S. boydii* membrane proteins were quite low. It also shows from the SDS-PAGE that OmpC is not visible in the *E. coli* extraction, which indicates that the solubilisation was not completely successful.

A more surprising result of the protein analysis was the identification of OprF as an outer membrane protein in Iso10. OprF is a porin which has earlier only been found in the *Pseudomonas* genus with one exception *(Azotobacter vinelandii)* (Rediers et al. 2004). It has been well studied because it is considered a potential vaccine protein for *Pseudomonas*. It also plays a role in antimicrobial drug resistance and functions as a porin. OprF shares a considerable sequence similarity to OmpA (56%) in the C-terminal end and is therefore considered as a member of the OmpA superfamily of porins (Bodilis and Barray 2006). The sequence from three CAF-fragments was identified with deNovo sequencing but two of them were overlapping giving only two sequences for database-search. The first CAF-fragment (1539.6**)** showed 100% identity (10/10) with the OprF sequence from *Pseudomonas* spp. The other CAF fragment (2335.0) had 100% identity matches (9/9) with both OprF from *Pseudomonas* spp. and a OmpA family protein from *Burkholderia* spp., which may be due to the C-terminal similarities between OprF and OmpA. Until now, no outer membrane analysis of *E. fergusonii* has been carried out and the pathogenic potential of *E. fergusonii* is still unclear. However, no earlier reports has indicated a connection between virulence factors in *Pseudomonas* and *E. fergusonii* and further analysis is needed to verify whether Iso10 really possesses the OprF gene and if the protein has the same functions in *E. fergusonii* as in *Pseudomonas.*

Three cytoplasmic proteins were identified (MetQ, GroEL/MopA and AtpD). Two of these proteins, GroEL and AtpD, has earlier been identified in membrane extractions by the carbonate method and are both reported as immunogenic proteins in *S. flexneri* (Ying et al 2005). GroEL is a very abundant protein and has been described as immunogenic in other bacteria such as *Chlamydia trachomatis* and *Heliobacter pylori*. Due to this, Ying et al (2005)

proposed that it may be a cross–reactive antigen from other bacterial infections and not unique for *S. flexneri.* However, AtpD is in the same publication described as a novel antigen. This protein lays at the inner the surface of the cytoplasmic membrane and is often categorized as an inner membrane protein. These proteins may also be immunogenic in Iso10, but this must be confirmed with Western blotting.

4.5 Discussion of results; Lipopolysaccharides

LPS from Iso10, *S. boydii* and *E. fergusonii* were extracted using a LPS extraction kit and analysed by SDS-PAGE and silver staining. A parallel gel was blotted on a membrane and probed with *S. boydii* 15- specific antiserum. The results clearly showed that Iso10 and *S. boydii* LPS cross- react, a feature not shared between the *E. fergusonii* type strain and *S. boydii.* These results confirm the earlier data by Rahman et al from the University of Dhaka which used whole cell extract to detect cross- reaction between Iso10 and *S. boydii*. They also observed smear of bands indicating a high similarity in O-antigenic properties(Rahman et al. 2007).

Complete serotyping is a complicated process involving a range of O- , capsular (K)- and flagellar (H) antisera which are only available in a few international reference laboratories. Due to this an alternative method for serotyping involving amplification of the *rfb*-cluster and characterisation by endonuclease restriction has been developed. The *rfb*-cluster encodes the enzymes involved in O- antigenic synthesis and is situated at position 44-45 minutes at the chromosome (Coimbra et al. 2000). Restriction analysis of the O- antigen gene cluster of Iso10 was previously performed by Nafisa Azmunda (Figure 1-4) and included in her Master thesis(Azmunda 2006). The results showed a significant O-antigen gene cluster homology between Iso10 and *S. boydii* 15 (Figure 1-8).

The *rfb* cluster plays an important role for the O-antigen variation in the immune response and due to this the selection pressure on the different O-antigenic forms is huge. The number of genes in the *rfb* cluster varies from 6 to 19 depending on the complexity of the polysaccharide chains. Closely related strains of *E. coli* possess different *rfb* genes giving different serotypes and distantly related strains may share identical *rfb* genes expressing the same O-antigen. Common O- antigenic pattern is also shared between *Shigella* and *E. coli* serotypes (Figure 4-3). Recombination of the *rfb* genes and the closely linked housekeeping gene *gnd* (encoding 6-phosphogluconate dehydrogenase which is cotransferred with *rfb*) is so frequent that the evolutionary relationships among strains acoording to LPS comparison are

not comparable with those proposed by MLEE and DNA sequencing. This strongly indicates that the O-antigenic variation among *E. coli* (including *Shigella*) species is by a large extent caused by horizontal gene transfer (Tarr et al. 2000). Numerous reports have also shown that even relatively distantly related bacterial strains can cross- react with antibodies specific for *E. coli* O157:H7, including *E. hermanii, Salmonella* group N, *Citrobacter freundii, Brucella melitensis* and *Yersinia enterocolitica* (Fegan et al. 2006).

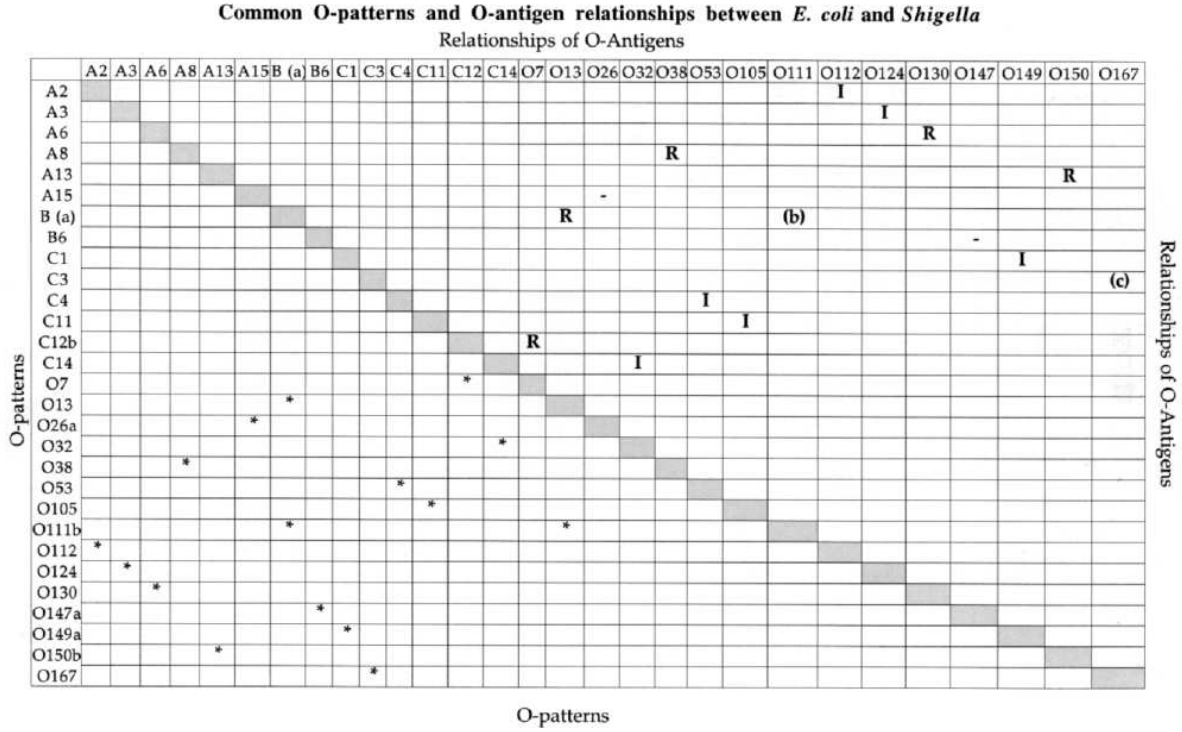


Figure 4-3: Pattern- sharing and serological cross-reaction between *Shigella* and *E. coli*.
: Identical pattern, I: Identical antigens, R: reciprocal cross- reaction, - : no cross- reaction known. a) agglutinates O111, b) serum anti-b c) serum anti –C3 agglutinates. A , B and C correspond to *S. dysenteriae* (A), *S. flexneri* (B), *S. boydii* (C), *respectively (Coimbra et al. 2000).

A phylogenetic tree including 87 enteric bacteria showed that 26 representatives of *E. coli*, 4 representatives of *Shigella* and *E. fergusonii* are closely related based on sequences of the *gnd* gene (Nelson and Selander 1994) (Figure 4-4). This indicates that genetic transfer of the *rfb-gnd* cluster has occurred among these strains.

The LPS cross- reaction between Iso10 and *S. boydii* in my results and the restriction analysis of the O antigen gene cluster strongly indicates a previous horizontal transfer of the *rfb*-cluster between *S. bodyii* and Iso10. No cross- reaction could be observed between the LPS from *E. fergusonii* type strain and *S. boydii*.

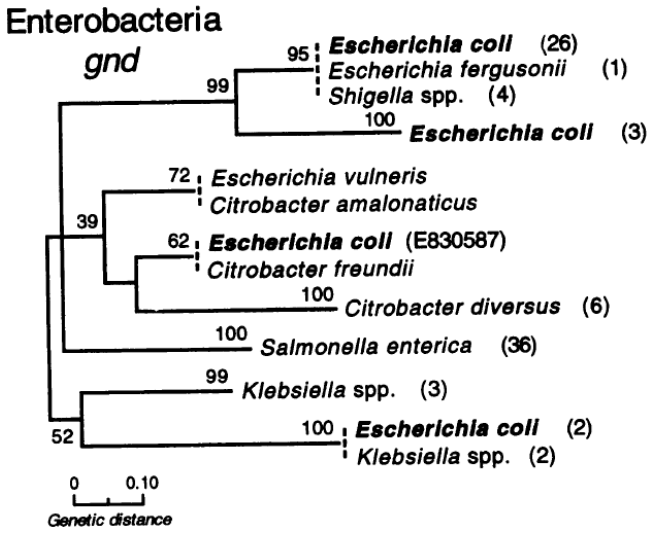


Figure 4-4: Part of a neighbour joining tree with 87 enteric bacteria based on sequences from the *gnd* gene which is closely linked to the O antigen gene cluster and often co-transported during gene transfer. Numbers of strains are indicated in parentheses. Bootstrap values are indicated in the nodes. Adopted from Nelson and Selander 1994.

These observations correlate with a previous study which reported a serological cross-reaction between an isolate of *E. fergusonii* and *E. coli* O157:H7 recovered during routine screening of beef (Fegan et al. 2006). They detected an identical gene order and high level of sequence identity between the *E. coli* O157:H7 *rfb* gene cluster and concluded that a horizontal gene transfer of the antigen gene cluster had occurred. My results, together with earlier observations suggest that *E. fergusonii* might receive new genes trough gene transfer from other members of the *E. coli/Shigella* group and obtain new pathogenic properties.

5 Conclusions and suggestions for future work

5.1 Conclusions

The aim of this study was to compare the environmental isolate Iso10 with the *S. boydii* 15 type strain based on multi locus sequence analysis, biochemical analysis and proteomic and immunological techniques.

The results from the multi locus sequence analysis of seven housekeeping genes and biochemical analysis showed Iso10 belongs to the *E. fergusonii* species.

Membrane proteins from Iso10 were separated with 2D electrophoresis and ten proteins were identified by deNovo sequencing (MALDI TOF-TOF MS) and Peptide Mass Fingerprinting with MALDI-TOF MS. One protein, OmpA was identified as immunogenic.

Western blotting of LPS extracted from Iso10 and *S. boydii* cross-reacted with *Shigella*-specific antisera. LPS from the *E. fergusonii* type strain did not cross-react.

The results indicate that the cross – reactivity between Iso10 and *S.boydii* is primarily caused by a horizontal gene transfer of the *rfb* O-antigen gene cluster. Outer membrane proteins in the *Escherichia* genus are so conserved that they might cross- react in Western blotting even though no gene transfer of the OMP genes have occurred, as showed in this study with OmpA.

If further analyses confirm that Iso10 possesses the *Pseudomonas* OprF porin gene it might indentify an exchange of antigens between species which has not earlier been observed.

Since Iso10 possesses a similar O- antigen gene cluster as *S. boydii* 15 and many of the OMPs are highly conserved among *Escherichia* members, Iso10 may induce an immune response sufficient to protect an organism against new invasions from *S. boydii* 15. However, my results strongly indicate that Iso10 is an *E. fergusonii* strain and may possess a range of virulence factors not yet identified. This strain thus has to be further characterized in order to evaluate it as a possible live shigellosis vaccine.

5.2 Suggestions for future work

- testing of virulence properties of Iso10 using animal models.
- Sequencing of the O antigen gene cluster (*rfb*) and the *gnd* housekeeping gene from Iso10 in order to verify horisontal gene transfer from *Shigella.*
- Comparison of the predicted membrane protein sequences with nucleotide sequences from the whole genome project of *E. fergusonii*.
- Verification of the presence of the O*prF* gene in Iso10 using Southern analysis or PCR.
- Complete genome sequencing of Iso10.
- If Iso10 lacks virulence properties/genes vaccination trials in animal models should be carried out.

List of abbreviations

°C	Celsius
µg	Microgram
µl	Microlitre
2D	Two dimensional
2DE	Two- dimensional electrophoresis
APS	Ammonium persulfate
ASB-14	Amidosulfobetaine-14
ATCC	American type culture collection
BLAST	Basic Local Aligment Search Tool
Bp	Basepair
CBB	Coomassie Brilliant Blue
Cfu	Colony forming unit
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
ddNTP	Dideoxyribonucleotide triphosphate
dH2O	Distilled water
DIG	Digoxigenin
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
dsDNA	Double stranded DNA
DTT	Dithiothreitol
DTT	ditiotreitol
e.g.	Exempli gratia, for example
EDTA	Ethylene diamine tetra-acetic acid
EMBL-EBI	European Molecular Biology Laboratory - European Bioinformatics Institute
Et al	And others
g	Gram
h	Hour
IEF	Isoelectric focusing
Kb	Kilobase
kDA	Kilo Dalton
L	Litre
LA	LB with agar
LB	Luria/Bertani- media
M	Molar
mA	Milliamp
min	Minutes
mM	Millimolar
mRNA	Messenger RNA
NCBI	Natural Center for Biotechnology Information
NJ	Neighbour- joining
OMP	Outer membrane protein
PBS	Phosphate buffer Saline
PCR	Polymerase chain Reaction
PMF	Peptide Mass Fingerprinting
PMSF	Phenylmethanesulphonylfluoride
RNA	Ribonueclicacid
Rpm	Potations per minute
rRNA	Ribosomal RNA
sdH2O	Sterile, distilled water

SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate- Polyacrylamide- gel electrophoresis
spp	Species
ssDNA	Single stranded DNA
TAE	Tris/acetate/EDTA- electrophoresis buffer
TBP	Tributyl phosphine
TBS	Tris borate saline- electrophoresis buffer
TCA	TriChloroAcetic
TGS	Tris/glycine/SDS- elecrophoresis buffer
Tm	Melting temperature
TOF-MS	Time Of Flight – Mass spectrometry
Tris	Trisaminemethane
UV	Ultraviolet light
v/v	Volume/volume
w/v	Weight/volume
WU-BLAST	Washington University BLAST

References

Arreaza L, Alcala B, Salcedo C, de la Fuente L, Vazquez JA (2003) Dynamics of the penA gene in serogroup C *meningococcal* strains. J Infect Dis 187:1010-1014

Azmunda N (2006) An immunomolecular Approach for the Identification of A typical *Shigella boydii* Isolated from Aquantic Ecosystem. In: Department of Microbiology. University of Dhaka, Dhaka, p 94

Bain MS, Green CC (1999) Isolation of *Escherichia fergusonii* in cases clinically suggestive of salmonellosis. Vet Rec 144:511

Bennish ML, Harris JR, Wojtyniak BJ, Struelens M (1990) Death in shigellosis: incidence and risk factors in hospitalized patients. J Infect Dis 161:500-506

Bodilis J, Barray S (2006) Molecular evolution of the major outer-membrane protein gene (oprF) of Pseudomonas. Microbiology 152:1075-1088

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248-254

Bruce Alberts AJ, Julian Lewis, Martin Raff, Keith Roberts, Peter Walter (2007) Chapter 10, Membrane structure. In: Molecular Biology of the Cell, Fifth Edition edn. Garland Science, Taylor &Francis group, New York

Coffey TJ et al. (1998) Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of *Streptococcus pneumoniae*. Mol Microbiol 27:73-83

Coimbra RS, Grimont F, Lenormand P, Burguiere P, Beutin L, Grimont PA (2000) Identification of *Escherichia coli* O-serogroups by restriction of the amplified O-antigen gene cluster (rfb-RFLP). Res Microbiol 151:639-654

Cole ST, Sonntag I, Henning U (1982) Cloning and expression in *Escherichia coli* K-12 of the genes for major outer membrane protein OmpA from *Shigella dysenteriae*, *Enterobacter aerogenes*, and *Serratia marcescens*. J Bacteriol 149:145-150

Cooper JE, Feil EJ (2004) Multilocus sequence typing--what is resolved? Trends Microbiol 12:373-377

Escobar-Paramo P, Giudicelli C, Parsot C, Denamur E (2003) The evolutionary history of Shigella and enteroinvasive *Escherichia coli* revised. J Mol Evol 57:140-148

Farmer JJ, 3rd et al. (1985) *Escherichia fergusonii* and *Enterobacter taylorae*, two new species of *Enterobacteriaceae* isolated from clinical specimens. J Clin Microbiol 21:77-81

Faruque SM et al. (2002) Isolation of *Shigella dysenteriae* type 1 and *S. flexneri* strains from surface waters in Bangladesh: comparative molecular analysis of environmental *Shigella* isolates versus clinical strains. Appl Environ Microbiol 68:3908-3913

Fegan N, Barlow RS, Gobius KS (2006) *Escherichia coli* O157 Somatic Antigen Is Present in an Isolate of *E. fergusonii*. Curr Microbiol

Formal SB, Hornick RB (1978) Invasive *Escherichia coli*. J Infect Dis 137:641-644

Fujiki Y, Hubbard AL, Fowler S, Lazarow PB (1982) Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum. J Cell Biol 93:97-102

Funke G, Hany A, Altwegg M (1993) Isolation of *Escherichia fergusonii* from four different sites in a patient with pancreatic carcinoma and cholangiosepsis. J Clin Microbiol 31:2201-2203

GEHealthcare (2004) 2D electrophoresis : Principles and Methods, third edition.

Gorg A, Postel W, Gunther S (1988) The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 9:531-546

Gorg A, Weiss W, Dunn MJ (2004) Current two-dimensional electrophoresis technology for proteomics. Proteomics 4:3665-3685
Guo H et al. (2005) Molecular analysis of the O-antigen gene cluster of *Escherichia coli* O86:B7 and characterization of the chain length determinant gene (wzz). Appl Environ Microbiol 71:7995-8001
Gygi SP, Rist B, Griffin TJ, Eng J, Aebersold R (2002) Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags. J Proteome Res 1:47-54
Hale TL (1991) Genetic basis of virulence in *Shigella* species. Microbiol Rev 55:206-224
Herzer PJ, Inouye S, Inouye M, Whittam TS (1990) Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. J Bacteriol 172:6175-6181
Jennison AV, Raqib R, Verma NK (2006) Immunoproteome analysis of soluble and membrane proteins of *Shigella flexneri* 2457T. World J Gastroenterol 12:6683-6688
Keusch GT, Jacewicz M (1977) The pathogenesis of *Shigella* diarrhea. VI. Toxin and antitoxin in *Shigella flexneri* and *Shigella sonnei* infections in humans. J Infect Dis 135:552-556
Koebnik R, Locher KP, Van Gelder P (2000) Structure and function of bacterial outer membrane proteins: barrels in a nutshell. Mol Microbiol 37:239-253
Kotloff KL et al. (1999) Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. Bull World Health Organ 77:651-666
Kweon MN (2008) Shigellosis: the current status of vaccine development. Curr Opin Infect Dis 21:313-318
Labrec EH, Schneider H, Magnani TJ, Formal SB (1964) Epithelial Cell Penetration as an Essential Step in the Pathogenesis of Bacillary Dysentery. J Bacteriol 88:1503-1518
Lan R, Alles MC, Donohoe K, Martinez MB, Reeves PR (2004) Molecular evolutionary relationships of enteroinvasive *Escherichia coli* and *Shigella* spp. Infect Immun 72:5080-5088
Larkin MA et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947-2948
Lawrence JG, Ochman H, Hartl DL (1991) Molecular and evolutionary relationships among enteric bacteria. J Gen Microbiol 137:1911-1921
Lopez JL (2007) Two-dimensional electrophoresis in proteome expression analysis. J Chromatogr B Analyt Technol Biomed Life Sci 849:190-202
Mahapatra A, Mahapatra S (2005) *Escherichia fergusonii*: an emerging pathogen in South Orissa. Indian J Med Microbiol 23:204
Maiden MC et al. (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A 95:3140-3145
Manz A, Pamme, N, Iossifidis D. (2004) Chapter 4, Mass Spectrometry. In: Bioanalytical Chemistry. Imperial College Presss, London
Maurelli AT, Sansonetti PJ (1988) Genetic determinants of *Shigella* pathogenicity. Annu Rev Microbiol 42:127-150
Mel DM, Terzin AL, Vuksic L (1965) Studies on vaccination against bacillary dysentery. 3. Effective oral immunization against *Shigella flexneri* 2a in a field trial. Bull World Health Organ 32:647-655
Melles DC et al. (2007) Comparison of multilocus sequence typing (MLST), pulsed-field gel electrophoresis (PFGE), and amplified fragment length polymorphism (AFLP) for genetic typing of *Staphylococcus aureus*. J Microbiol Methods 69:371-375

Molloy MP (2000) Two-dimensional electrophoresis of membrane proteins using immobilized pH gradients. Anal Biochem 280:1-10
Molloy MP et al. (2000) Proteomic analysis of the *Escherichia coli* outer membrane. Eur J Biochem 267:2871-2881
Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol 51 Pt 1:263-273
Nelson K, Selander RK (1994) Intergeneric transfer and recombination of the 6-phosphogluconate dehydrogenase gene (gnd) in enteric bacteria. Proc Natl Acad Sci U S A 91:10227-10231
Niyogi SK (2007) Increasing antimicrobial resistance--an emerging problem in the treatment of shigellosis. Clin Microbiol Infect 13:1141-1143
Ochman H, Selander RK (1984) Standard reference strains of *Escherichia coli* from natural populations. J Bacteriol 157:690-693
Ochman H, Whittam TS, Caugant DA, Selander RK (1983) Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*. J Gen Microbiol 129:2715-2726
O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. J Biol Chem 250:4007-4021
Paradis S et al. (2005) Phylogeny of the *Enterobacteriaceae* based on genes encoding elongation factor Tu and F-ATPase beta-subunit. Int J Syst Evol Microbiol 55:2013-2025
Passwell JH et al. (2001) Safety and immunogenicity of improved *Shigella* O-specific polysaccharide-protein conjugate vaccines in adults in Israel. Infect Immun 69:1351-1357
Prescott LM, Harley, John P and Klein, Donald A. (2005) In: Microbiology, 6th edn. McGraw-Hill, New York, pp 56-59
Pupo GM, Karaolis DK, Lan R, Reeves PR (1997) Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and mdh sequence studies. Infect Immun 65:2685-2692
Pupo GM, Lan R, Reeves PR (2000) Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. Proc Natl Acad Sci U S A 97:10567-10572
Raetz CR, Whitfield C (2002) Lipopolysaccharide endotoxins. Annu Rev Biochem 71:635-700
Rahman MZ, Sultana M, Khan SI, Birkeland NK (2007) Serological cross-reactivity of environmental isolates of *Enterobacter, Escherichia, Stenotrophomonas,* and *Aerococcus* with *Shigella* spp.-specific antisera. Curr Microbiol 54:63-67
Rediers H, Vanderleyden J, De Mot R (2004) Azotobacter vinelandii: a *Pseudomonas* in disguise? Microbiology 150:1117-1119
Roy S, Das AB, Biswas T (1994a) Antigenicity and antigenic relatedness of the outer membrane proteins of Shigella species. Microbios 79:55-64
Roy S, Das AB, Ghosh AN, Biswas T (1994b) Purification, pore-forming ability, and antigenic relatedness of the major outer membrane protein of *Shigella dysenteriae* type 1. Infect Immun 62:4333-4338
Sanger F, Nicklen S, Coulson AR (1992) DNA sequencing with chain-terminating inhibitors. 1977. Biotechnology 24:104-108
Sansonetti PJ, d'Hauteville H, Ecobichon C, Pourcel C (1983) Molecular comparison of virulence plasmids in *Shigella* and enteroinvasive *Escherichia coli*. Ann Microbiol (Paris) 134A:295-318

Sansonetti PJ, Kopecko DJ, Formal SB (1981) *Shigella sonnei* plasmids: evidence that a large plasmid is necessary for virulence. Infect Immun 34:75-83
Schnaitman CA, Klena JD (1993) Genetics of lipopolysaccharide biosynthesis in enteric bacteria. Microbiol Rev 57:655-682
Smith SG, Mahon V, Lambert MA, Fagan RP (2007) A molecular Swiss army knife: OmpA structure, function and expression. FEMS Microbiol Lett 273:1-11
Suckau D, Resemann A, Schuerenberg M, Hufnagel P, Franzen J, Holle A (2003) A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. Anal Bioanal Chem 376:952-965
Tarr PI, Schoening LM, Yea YL, Ward TR, Jelacic S, Whittam TS (2000) Acquisition of the rfb-gnd cluster in evolution of *Escherichia coli* O55 and O157. J Bacteriol 182:6183-6191
Urwin R, Maiden MC (2003) Multi-locus sequence typing: a tool for global epidemiology. Trends Microbiol 11:479-487
Venkatesan M, Fernandez-Prada C, Buysse JM, Formal SB, Hale TL (1991) Virulence phenotype and genetic characteristics of the T32-ISTRATI *Shigella flexneri* 2a vaccine strain. Vaccine 9:358-363
von Seidlein L et al. (2006) A multicentre study of *Shigella* diarrhoea in six Asian countries: disease burden, clinical manifestations, and microbiology. PLoS Med 3:e353
Weiner JH, Li L (2008) Proteome of the *Escherichia coli* envelope and technological challenges in membrane proteome analysis. Biochim Biophys Acta 1778:1698-1713
WHO (2005) Guidelines for the control of shigellosis, including epidemics due to *Shigella dysenteriae* type 1 (http://www.who.int/child_adolescent_health/documents/diarrhoea/en/index.html) Acess date:1 September 2008, ISBM: 9241592330. In:
WHO (Updated February 2009) Diharrhoeal diseases: Shigellosis, Acess date: 15.Feb.2009 (http://www.who.int/vaccine_research/diseases/diarrhoeal/en/index6.html). In:
Wilkins MR, Gasteiger E, Sanchez JC, Bairoch A, Hochstrasser DF (1998) Two-dimensional gel electrophoresis for proteome projects: the effects of protein hydrophobicity and copy number. Electrophoresis 19:1501-1505
Wirth T et al. (2006) Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol Microbiol 60:1136-1151
Wragg P et al. (2008) Characterisation of *Escherichia fergusonii* isolates from farm animals using an *Escherichia coli* virulence gene array and tissue culture adherence assays. Res Vet Sci
Yang J et al. (2007) Revisiting the molecular evolutionary history of *Shigella* spp. J Mol Evol 64:71-79
Ying TY et al. (2005) Immunoproteomics of membrane proteins of *Shigella flexneri* 2a 2457T. World J Gastroenterol 11:6880-6883

APPENDIX A

Media

LB Broth

10g tryptone

5 g yeast extract

10g NaCl

dH_2O to 1000ml

Kits

GenElute ™ Bacterial genomic DNA kit (SIGMA)

GenElute™ PCR clean–up kit (SIGMA)

Gel extraction kit (SIGMA)

Big Dye Terminator v3.1. Cycle sequencing kit (Quiagene)

Ettan ™ Sample preparations kits and reagents, 2D clean up kit (Amersham biosciences)

API 20 E (Biomérieux)

LPS Extraction kit (iNtRON Biotechnology)

SilverQuest ™ Silver staining kit (Invitrogen)

APPENDIX B

Buffers and solutions for protein extraction and fractionation

Washing Buffer(2DE)

8 mM NaCl

2mM KCl

1.5 mM KH_2PO_4

9mM NaH_2PO_4

Lysis buffer(2DE)

8mol/L urea

4% W/v chaps

1% W/v DTT

0.8 % ampholytes 3–7

35mmol/L TrisHCl

5mM EDTA

1mM PMSF

Coomassie brilliant blue (CBB) solution(2DE)

150ml coomassie brilliant blue G–250

150ml dH_2O

200ml H_2PO4

Rehydration buffer without bromophenol blue (2DE)

7M Urea

2M Thiourea

4% W/v CHAPS

20 mM DTT

0.5% Triton X –100

0.5% v/v pharmalyte 4–7

Stored in 2 ml ependorf tubes at –80°C

Rehydration buffer with bromophenol blue (2DE)

Prepeared as rehydration buffer without bromophenol blue, in addition a few grains of bromophenol blue is added. Stored in 2ml ependorf tubes at –80°C.

Membrane solubilization solution 1(2DE)

7M Urea

2M Thiourea

2mM TBP

20mM Tris– base

1% ASB14

0.5% 4–7 Biolythes

Membrane solubilization solution 2(2DE)

7M Urea

2M Thiourea

2% ASB14

1% DTT

0.5 % Ampolytes

SDS equilibration buffer(2DE)

50mM Tris HCl pH 8,8

6M urea

30% (v/v) Glycerol

2% (w/v) SDS

A few grains of bromophenol blue

SDS equilibration buffer with DTT(2DE)

2.5 mg DTT /ml equilibration buffer

DTT is added prior to use

SDS equilibration buffer with iodacetamide(2DE)

45mg iodacetamide/ml equilibration buffer

Iodacetamide is added prior to use.

5x Running buffer(2DE)

25mM Tris base

192mM Glycin

0.1%(w/v) SDS

Stored at room temperature. Diluted to 1 x Running buffer before use.

Fixation solution(2DE):

28ml 85% Phosphoric acid

dH_2O to 2000ml

Solution 1(2DE):

680 ml Methanol

340 g ammonium sulphate

28ml 85% Phosphoric acid

dH_2O to 2000ml

Ammonuim sulohate is solubilized in 500ml dH_2O and phosphoric acid. Methanol is mixed with aqua dest and slowly added to the solution otherwise it will precipitate.

Staining solution(2DE)

Like solution 1 with 1.32 g CBB G250

dH_2O to 2000ml

CBB is solubilized in methanol before added to the solution.

Gel storage solution(2DE)

5% Sodium acid.

SDS–PAGE sample buffer

20ml dH_2O

5ml 0.5MTris HCl pH 6,8

4ml Glycerol

8ml 10% (w/v) SDS

1ml 0.05% (w/v) Bromophenolblue

ß–mercaptoetanol 1:4 relationship

10 x TGS–Buffer pH 8,3

18 g Tris base

86.4g Glycine

6g SDS

dH_2O to 600ml

APPENDIX C
Buffers and solutions for Western blotting

Anodic solution 1
0.3M Tris HCl

20% (v/v) methanol

pH adjusted to 10,4

Anodic solution 2
25mM Tris HCl

20% (v/v) methanol

pH adjusted to 10.4

Cathodic solution
4mM 6 amino–n–hexanoic acid

20% (v/v) methanol

pH adjusted to 7.6

10 x Blotting buffer (Stock solution)
7.57g Tris

36.0 g Glycin

dH2O to 250 ml

1 x Blotting buffer
100 ml 10 x Blotting buffer

200ml Methanol

700 ml dH2O

10 x TBS buffer
200mM Tris

5M Nacl

dH2O to 0.5 l

pH adjusted to 7.5

0,1% Tween 20 TBS buffer
0.2 ml Tween 20

200 ml TBS buffer

APPENDIX D
Alternative reaction mix and programme for Phusion Polymerase

Components	**Volume added (µl)**
Sd H2O	35.5
5 x PCR reactionbuffer	10
dNTP (10mM of each)	1
Primer F (25pmol/µl)	1
Primer R (25pmol/µl)	1
Polymerase, Phusion (2units/lµl)	0.5
DNA, template (25ng/µl)	1
Total volume	**50**

PCR programme:

98°C x 3min (98°C x 15sec – annealing temperature x 30sec – 72°C x 40sec) x 30, 72°C x 10 min – 4°C ∞

APPENDIX E
Blast search with complete sequences

Results from EMBL–EBI blast search with the complete sequences from nine housekeeping genes.

Gene	Acc.number	Organism	Length (bp)	Identity %	E- value
Adk	AY686533	Escherichia fergusonii	536	99	6.3e-111
	BA000007	Escherichia coli O157:H7 str. Sakai DNA	5498450	93	9.3e-102
	AE005174	Escherichia coli O157:H7 EDL933	5528445	93	9.3e-102
	X03038	E. coli	1055	92	2.5e-101
	CP000036	Shigella boydii Sb227,	4519823	93	6.1e-101
FumC	X04065	Escherichia coli (4.2.1.2.)	2250	83	2.0e-110
	EA375705	Sequence 24528 from unidentified patent US 7314974.	1404	83	3.2e-110
	U00096	Escherichia coli K12 MG1655	4639675	83	6.8e-109
	DD367313	Escherichia coli W3110	4641433	83	6.8e-109
	AP009048	Escherichia coli W3110	4646332	83	6.8e-109
GyrB	X04341	Escherichia coli K12	4931	97	1.0e-160
	D87842	Escherichia coli K12	2415	97	2.1e-160
	CS362320	Escherichia coli Patent WO2004108933.	2412	97	2.1e-160
	L10328	Escherichia coli	136254	97	7.2e-159
	CP000036	Shigella boydii Sb227	4519823	97	7.6e-159
Icd	AY132849	Escherichia fergusonii strain ATCC 35469	1166	99	1.0e-174
	AF017601	Escherichia coli EC40	1212	97	3.2e-170
	AE014073	Shigella flexneri 2a str. 2457T	4599354	97	1.3e-169
	AE005674	Shigella flexneri 2a str. 301	4607203	97	1.3e-169
	AF017598	Escherichia coli EC37	1212	96	5.2e-169
Mdh	AF091766	Escherichia coli, isolate RT019A	864	85	1.1e-130
	AF293136	Escherichia coli isolate M1340_B10	1541	85	1.5e-130
	AF293137	Escherichia coli isolate M1345_B6	1541	85	1.5e-130
	AF071028	Escherichia coli strain E2348/69	863	85	2.7e-130
	AF071030	Escherichia coli isolate C54-58	863	85	2.7e-130
PurA	CS191906	Escherichia coli, Sequence 150	1299	98	5.2e-166

		from Patent WO2005103073.			
	J04199	Escherichia coli	2726	98	6.3e-166
	A92595	Escherichia coli, Sequence 1 from Patent WO9810074	1311	98	1.3e-165
	BD006127	Escherichia coli	1311	98	1.3e-165
	EA398532	Unidentified, Sequence 47356 from patent US 7314974.	1299	98	1.3e-165
RecA	BD064136	Synthetic construct; artificial sequences	1398	98	5.4e-117
	BD233595	Synthetic construct; artificial sequences	1398	98	5.4e-117
	CS423861	Synthetic construct; artificial sequences. Sequence 14 from Patent EP1707641	1398	98	5.4e-117
	CS443102	Synthetic construct; artificial sequences. Sequence 7 from Patent EP1717322	1398	98	5.4e-117
	AR160050	Synthetic construct; artificial sequences. Sequence 13 from patent US 6251674.	1398	98	5.4e-117
trpB	AF330405	Shigella dysenteriae strain SD0292	1138	99	3.4e-242
	AF330406	Shigella dysenteriae strain SD0297	1138	99	3.4e-242
	AF330392	Shigella boydii strain SB1156	1138	99	3.7e-240
	U25418	Escherichia coli ECOR 15	5260	97	9.3e-238
	U25422	Escherichia coli ECOR 24	5260	97	9.3e-238
PurN	AF293166	Escherichia coli isolate ECOR37	2101	80	5.4e-136
	AF293165	Escherichia coli isolate M1372_D1	2101	80	8.9e-135
	AF293163	Escherichia coli isolate M1367_D10	2101	80	8.9e-135
	AF293210	Escherichia coli isolate M1371_D6	2101	80	1.4e-134
	:AF293161	Escherichia coli isolate ECOR40	2101	80	2.3e-134

APPENDIX F
Aligments

Aligment of Iso10 and *Shigella* spp. – 16S rRNA sequences

```
Shigella sp. BBDP80 16S ribosomal RNA gene, partial
            sequence.
 Length = 1503
 Plus Strand HSPs:
 Score = 7121 (1074.5 bits), Expect = 0., P = 0.
 Identities = 1427/1429 (99%), Positives = 1427/1429 (99%), Strand = Plus / Plus

Query:     1 TGCA-GTCGAACGGTAACAGGAATCAGCTTGCTGATTCGCTGACGAGTGGCGGACGGGTG 59
             |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    49 TGCAAGTCGAACGGTAACAGGAATCAGCTTGCTGATTCGCTGACGAGTGGCGGACGGGTG 108

Query:    60 AGTAATGTCTGGGAAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATAC 119
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   109 AGTAATGTCTGGGAAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATAC 168

Query:   120 CGCATAATGTCGCAAGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATCGGATGTGCCC 179
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   169 CGCATAATGTCGCAAGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATCGGATGTGCCC 228

Query:   180 AGATGGGATTAGCTTGTAGGTGGGGTAACGGCTCACCTAGGCGACGATCCCTAGCTGGTC 239
             |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   229 AGATGGGATTAGCTAGTAGGTGGGGTAACGGCTCACCTAGGCGACGATCCCTAGCTGGTC 288

Query:   240 TGAGAGGATGACCAGCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGC 299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   289 TGAGAGGATGACCAGCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGC 348

Query:   300 AGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAA 359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   349 AGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAA 408

Query:   360 GGCCTTCGGGTTGTAAAGTACTTTCAGCGGGGAGGAAGGGAGTAAAGTTAATACCTTTGC 419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   409 GGCCTTCGGGTTGTAAAGTACTTTCAGCGGGGAGGAAGGGAGTAAAGTTAATACCTTTGC 468

Query:   420 TCATTGACGTTACCCGCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATA 479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   469 TCATTGACGTTACCCGCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATA 528

Query:   480 CGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTA 539
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   529 CGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTA 588

Query:   540 AGTCAGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGA 599
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   589 AGTCAGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGA 648

Query:   600 GTCTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGA 659
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   649 GTCTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGA 708

Query:   660 ATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGG 719
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   709 ATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGG 768

Query:   720 GAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGTCGACTTGGAGGTT 779
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   769 GAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGTCGACTTGGAGGTT 828

Query:   780 GTGCCCTTGAGGCGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAGTACGG 839
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct:   829 GTGCCCTTGAGGCGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAGTACGG 888

Query:   840 CCGCAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGT 899
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   889 CCGCAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGT 948

Query:   900 TTAATTCGATGCAACGCGAAGAACCTTACCTGGTCTTGACATCCACGGAAGTTTTCAGAG 959
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   949 TTAATTCGATGCAACGCGAAGAACCTTACCTGGTCTTGACATCCACGGAAGTTTTCAGAG 1008

Query:   960 ATGAGAATGTGCCTTCGGGAACCGTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGT 1019
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1009 ATGAGAATGTGCCTTCGGGAACCGTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGT 1068

Query:  1020 TGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTTGCCAGCGGT 1079
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1069 TGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTTGCCAGCGGT 1128

Query:  1080 CCGGCCGGGAACTCAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTC 1139
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1129 CCGGCCGGGAACTCAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTC 1188

Query:  1140 AAGTCATCATGGCCCTTACGACCAGGGCTACACACGTGCTACAATGGCGCATACAAAGAG 1199
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1189 AAGTCATCATGGCCCTTACGACCAGGGCTACACACGTGCTACAATGGCGCATACAAAGAG 1248

Query:  1200 AAGCGACCTCGCGAGAGCAAGCGGACCTCATAAAGTGCGTCGTAGTCCGGATTGGAGTCT 1259
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1249 AAGCGACCTCGCGAGAGCAAGCGGACCTCATAAAGTGCGTCGTAGTCCGGATTGGAGTCT 1308

Query:  1260 GCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAA 1319
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1309 GCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAA 1368

Query:  1320 TACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCAAAAGAAG 1379
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1369 TACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCAAAAGAAG 1428

Query:  1380 TAGGTAGCTTAACCTTCGGGAGGGCGCTTACCACTTTGTGATTCATGAC 1428
             ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1429 TAGGTAGCTTAACCTTCGGGAGGGCGCTTACCACTTTGTGATTCATGAC 1477
```

Aligments of Iso10 and *E. fergusonii* ATCC 35469- 16S rRNA sequences

```
AF530475 Escherichia fergusonii 16S ribosomal RNA gene,partial sequence.
Length = 1473
Plus Strand HSPs:
Score = 7108 (1072.5 bits), Expect = 0., P = 0.
Identities = 1425/1429 (99%), Positives = 1426/1429 (99%), Strand = Plus / Plus

Query:     1 TGCA-GTCGAACGGTAACAGGAATCAGCTTGCTGATTCGCTGACGAGTGGCGGACGGGTG 59
             |||| |||||||||||||||+||||||||||||||||||||||||||||||||||||||
Sbjct:    22 TGCAAGTCGAACGGTAACAGGRATCAGCTTGCTGATTCGCTGACGAGTGGCGGACGGGTG 81

Query:    60 AGTAATGTCTGGGAAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATAC 119
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    82 AGTAATGTCTGGGAAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATAC 141

Query:   120 CGCATAATGTCGCAAGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATCGGATGTGCCC 179
             ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   142 CGCATAACGTCGCAAGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATCGGATGTGCCC 201

Query:   180 AGATGGGATTAGCTTGTAGGTGGGGTAACGGCTCACCTAGGCGACGATCCCTAGCTGGTC 239
             ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   202 AGATGGGATTAGCTAGTAGGTGGGGTAACGGCTCACCTAGGCGACGATCCCTAGCTGGTC 261

Query:   240 TGAGAGGATGACCAGCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGC 299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   262 TGAGAGGATGACCAGCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGC 321

Query:   300 AGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAA 359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   322 AGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAA 381

Query:   360 GGCCTTCGGGTTGTAAAGTACTTTCAGCGGGGAGGAAGGGAGTAAAGTTAATACCTTTGC 419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   382 GGCCTTCGGGTTGTAAAGTACTTTCAGCGGGGAGGAAGGGAGTAAAGTTAATACCTTTGC 441

Query:   420 TCATTGACGTTACCCGCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATA 479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   442 TCATTGACGTTACCCGCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATA 501

Query:   480 CGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTA 539
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   502 CGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTA 561

Query:   540 AGTCAGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGA 599
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   562 AGTCAGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGA 621

Query:   600 GTCTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGA 659
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   622 GTCTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGA 681

Query:   660 ATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGG 719
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   682 ATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGG 741

Query:   720 GAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGTCGACTTGGAGGTT 779
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   742 GAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGTCGACTTGGAGGTT 801

Query:   780 GTGCCCTTGAGGCGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAGTACGG 839
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   802 GTGCCCTTGAGGCGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAGTACGG 861

Query:   840 CCGCAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGT 899
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   862 CCGCAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGT 921

Query:   900 TTAATTCGATGCAACGCGAAGAACCTTACCTGGTCTTGACATCCACGGAAGTTTTCAGAG 959
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   922 TTAATTCGATGCAACGCGAAGAACCTTACCTGGTCTTGACATCCACGGAAGTTTTCAGAG 981

Query:   960 ATGAGAATGTGCCTTCGGGAACCGTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGT 1019
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   982 ATGAGAATGTGCCTTCGGGAACCGTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGT 1041

Query:  1020 TGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTTGCCAGCGGT 1079
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1042 TGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTTGCCAGCGGT 1101

Query:  1080 CCGGCCGGGAACTCAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTC 1139
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1102 CCGGCCGGGAACTCAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTC 1161

Query:  1140 AAGTCATCATGGCCCTTACGACCAGGGCTACACACGTGCTACAATGGCGCATACAAAGAG 1199
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1162 AAGTCATCATGGCCCTTACGACCAGGGCTACACACGTGCTACAATGGCGCATACAAAGAG 1221

Query:  1200 AAGCGACCTCGCGAGAGCAAGCGGACCTCATAAAGTGCGTCGTAGTCCGGATTGGAGTCT 1259
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1222 AAGCGACCTCGCGAGAGCAAGCGGACCTCATAAAGTGCGTCGTAGTCCGGATTGGAGTCT 1281

Query:  1260 GCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAA 1319
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1282 GCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAA 1341

Query:  1320 TACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCAAAAGAAG 1379
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1342 TACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCAAAAGAAG 1401

Query:  1380 TAGGTAGCTTAACCTTCGGGAGGGCGCTTACCACTTTGTGATTCATGAC 1428
            ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1402 TAGGTAGCTTAACCTTCGGGAGGGCGCTTACCACTTTGTGATTCATGAC 1450
```

APPENDIX G

PROTOCOLS

Protocol for production of antibodies (Dhaka University)

Extraction of proteins by water:

40 ml overnight culture is centrifuged at 10 000g for 10 min at 10 °C. The pellets are washed three times with normal seline. Finally the pellet is dissolved in 40ml of 1% SDS in sterile water and incubated over night in a shaker incubator at room temperature. The suspension is then centrifuged at 10 000g for 10 min at 10°C. The supernatant is filtered through 0.45μm Milipore membrane and stored at – 20° C.

Extraction of proteins by extraction buffer:

Bacterial pellets are obtained from culture in exponential phase. The pellet is washed twice with PBS (pH 7.4) and resuspended in extraction buffer (0.05M Tris pH 6.8, 2%SDS, 10% Glycerol) before it is boiled for 3 min. The culture supernatant is separated and filtered though 0.45 μm Milipore membrane and stored at 20°C.

Immunization of rabbits:

Eight New Zealand white rabbits (1,5 – 2,0 kg weight) with proteins from water extaction and extraction by extraction buffer. Within 10 days of the final dose, the antibody development was confirmed by slide agglutination test. The rabbits were bled by heart puncturing to collect fresh blood. The blood samples were incubated at 37 °C. The blood samples were then kept at 4 °C overnight and sera were separated. The sera were aliquoted in 1.5 ml eppendorf tubes and stored at –20°C.

PMF and deNovo sequencing protovol (PROBE)

PMF protocol (performed by PROBE)

Sample preparation for MALDI-ToF MS

1. Poros Dried Droplet (DD) preparation; concentration/desalting of samples;

Prepare microcolumns of Poros 20 R2 (Applied Biosystems, USA) in Gelloader tips (M. Kussmann et al. 1997, J. Mass Spec. 32: 593-601). Wet column by dispensing 10 µl 50% ACN (acetonitrile) in MilliQ water and condition it with 10 µl 1% TFA (trifluoroacetic acid). Bind sample (5-10 µl). Wash sample with 10 µl 0.1% TFA, and elute with 1 µl CHCA (a-cyano-4-hydroxycinnamic acid, 5-10 mg/ml 60% ACN/15% methanol/0.1% TFA) matrix solution directly on MALDI target as several small drops. Apply a peptide standard (peptide calibration standard from BRUKER Daltonic GmbH, Leipzig, Germany) dissolved in matrix as nearest neighbor.

2. µ-ZipTip DD preparation; concentration/desalting of samples;

Wet column by aspirating and dispensing (to waste) 10 µl 50% ACN in MilliQ water, repeat two times (**take care not to get air into column**). Condition column by aspirating and dispensing (to waste) 10 µl 0.1% TFA, repeat two times. Bind 5-10 µl sample (if less than 10 µl, dilute in 0.1% TFA) by slowly aspirating and dispensing sample 10 times through column. Wash sample by aspirating and dispensing (to waste) 10 µl 0.1% TFA, repeate two times. Dry column (aspirate and dispense air). Add 1 µl CHCA (see above) matrix solution into a 0.2 ml Eppendorf tube, and slowly aspirate and dispense this solution through the column 3 times. Apply directly on MALDI target from ZipTip. Make an DD- preparation of a peptide standard (see above) dissolved in matrix as nearest neighbor.

3. AnchorChip (600µm) DD preparation;

Apply 1-2 µl CHCA matrix (0.3 mg/ml in ethanol : aceton 2:1) onto the anchorChip target, and mix it with 0.2-1 µl analyte (sample solution may contain 0-70% organic solvents). Let sample dry. Wash sample as follows: Apply 5-10 µl 10 mM ammonium phosphate dissolved in 0.1% TFA on sample for a few seconds following removal with a pipette tip. Let sample dry. Recrystallize by applying 1 µl ethanol : aceton : 0.1% TFA (6:3:1), and let sample dry. At standard locations, apply 1 µl of a solution composed of 1µl peptide standard (see above) and 5µl CHCA matrix. Let sample dry (desiccator).'

Obtaining the peptide mass fingerprint

4. Mass spectrometric analysis;

Analyse samples using MALDI-ToF MS to produce the fingerprint mass spectrum. Calibrate mass spectra using the calibration curve obtained from a peptide standard spectrum (external calibration), i.e. calibration standard no. 206195 from BRUKER Daltonic GmbH (Leipzig, Germany). The peptides in this standard cover a mass range from 1046 - 3147 Da.

Protein identification

5. Using MS (ToF data) and MSMS (PSD data) for protein identification;

Search with obtained monoisotopic mass maps of tryptic peptides in protein databases (NCBI nr, MSDB, Swissprot) using the Mascot (www.matrixscience .com) or Profound (http://prowl.rockefeller.edu/profound_bin/WebProFound.exe) search engine. From BioTools (software from Bruker Daltonic GmbH, Germany) ToF-data and PSD data can be searches directly in the Mascot search engine, and this software also offer a combination search composed of the MS-mass map together with all PSD mass maps (MSMS spectra) of the same sample.

DeNovo sequencing of modified tryptic peptides (Chemically Assisted Fragmentation, CAF) using MALDI-ToF

Remove organic solvents (e.g. acetonitrile (ACN), methanol (MeOH)) from sample by vacuum drying in a Rotavapor (Concentrator 5301 from Eppendorf AG, Hamburg, Germany) till it remains 10-15 µl sample. If necessary, add 15 µl 0.1% TFA and vacuum dry to about the same volume as above.

Obtain preCAF spectrum of tryptic peptides

1. Poros Dried Droplet (DD) preparation; concentration/desalting of samples;

Prepare micro columns of Poros 20 R2 (Applied Biosystems, USA) in Gelloader tips (M. Kussmann et al. 1997, J. Mass Spec. 32: 593-601). Wet column by dispensing 10 µl 50% ACN in MilliQ water and condition it with 10 µl 1% TFA. Bind $^{1}/_{10}$ of sample onto the Poros column. Wash sample with 10 µl 0.1% TFA, and elute with 1 µl DHB (2,5-dihydroxybenzoic acid, 10-20 mg/ml 60% ACN/0.1% TFA) matrix solution directly on MALDI target as 2-3 small drops (don't need to elute all). Apply peptide standard (calibration standard from BRUKER Daltonic GmbH, Leipzig, Germany) dissolved in matrix as nearest neighbor.

2. µ-ZipTip DD preparation; concentration/desalting of samples;

Wet column by aspirating and dispensing (to waste) 10 µl 50% ACN in MilliQ water, repeat two times (take care not to get air into column). Condition column by aspirating and dispensing (to waste) 10 µl 0.1% TFA, repeat two times. Bind $^{1}/_{10}$ of sample (if less than 10 µl, dilute in 0.1% TFA) by slowly aspirating and dispensing sample 10 times through column. Wash sample by aspirating and dispensing (to waste) 10 µl 0.1% TFA, repeate two times. Dry column (aspirate and dispense air). Add 1 µl DHB (10-20 mg DHB/ml 60% ACN /0.1% TFA) matrix solution into a 0.2 ml Eppendorf tube, and slowly aspirate and dispense this solution through the column 3 times, and apply directly on MALDI target. Apply peptide standard (see above) dissolved in matrix as nearest neighbor.

3. AnchorChip DD preparation;

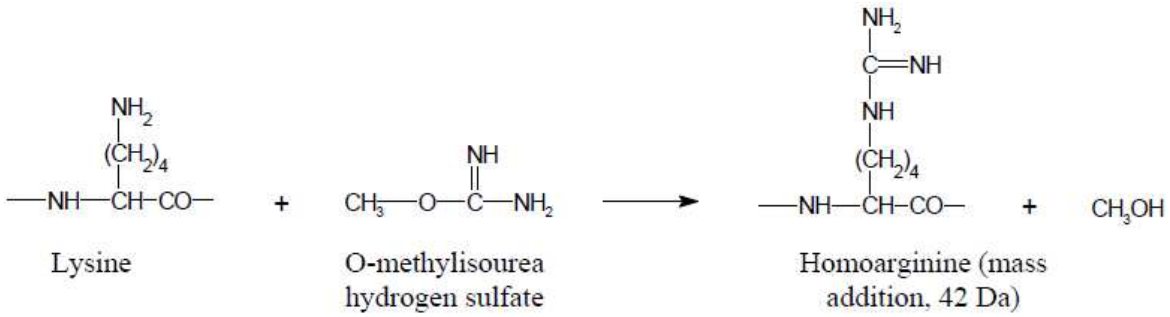
Apply 0.5-1 µl DHB matrix (5-10 mg/ml in 0.1% TFA : ACN 2:1) on a 400/600 µm anchorChip target (BRUKER Daltonic GmbH, Germany), and mix with 0.5-1 µl analyte solution. At standard locations, apply 1 µl of a solution composed of 1µl peptide standard (see above) and 5µl DHB matrix. Let sample dry (desiccator). Proteomic Unit at University of Bergen

4. <u>Analysis on MALDI ToF;</u>

Produce a mass fingerprint of the sample using MALDI-ToF MS. Calibrate samples using the calibration curve obtained from the peptide standard spectrum (external calibration), i.e. calibration standard no. 206195 from BRUKER Daltonic GmbH (Leipzig, Germany). The peptides in this standard cover a mass range from 1046 - 3147 Da.

<u>Obtain postCAF spectrum of tryptic peptides</u>

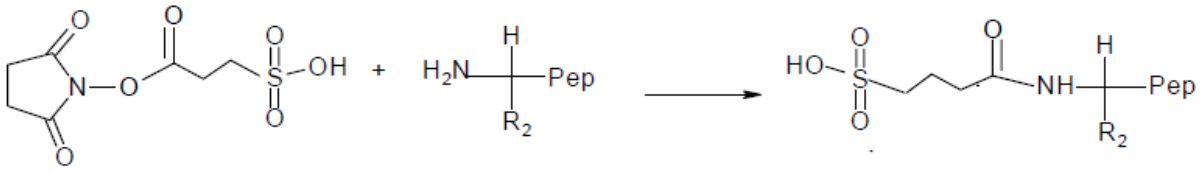
5. <u>Derivatization of lysine side chains;</u>



Lysine — O-methylisourea hydrogen sulfate — Homoarginine (mass addition, 42 Da)

Dissolve dry sample in 20 to 40ul of **DIEA-solution** (di-isopropylethylamine, H_2O : DIEA 19:1). If sample is already in solution, add 2ul concentrated DIEA, ***check to make sure solution is basic***. Add 2 to 4ul of **reagent solution** (86mg O- methylisourea hydrogensulphate /ml of 90:10 water/DIEA), and vortex well. **Check pH - if basic, ok** (pH 10 or a little above). If not, add 1ul concentrated DIEA at a time and check pH until basic. Use pH paper and look for a nice blue color. React overnight at room temperature. Check the pH again before leaving for the day, just to make sure it stays basic. The next morning, quench by adding 1M HCl until acidic, usually 5 to 20 ul.

This derivatization adds 42 Da to each Lysine residue, and is only necessary if doing DeNovo sequencing of Lys containing peptides.



NHS-ester of 3-Sulfo-propionic acid anhydride — The N-terminal of a peptide — Sulfonated peptide (mass addition, +136 Da)

Wet a C_{18} µ-ZipTip by aspirating and dispensing (to waste) 10µl 50% ACN in MilliQ water, repeat two times (take care not to get air into column). Condition column by aspirating and dispensing (to waste) 10 µl 0.1% TFA, repeat two times. Bind sample (if less than 10 µl, dilute in 0.1% TFA) by slowly aspirating and dispensing sample 10 times through column. Wash sample by aspirating and dispensing (to waste) 10 µl 0.1% TFA, repeate two times.

Dissolve immediately before use 0.6 mg CAF-reagent (store dry in freezer) in 6 µl CAF-buffer (0.25 M Sodium bicarbonate, pH 9.4) in a 0.2 ml Eppendorf tube (aspirate and dispense a few times to dissolve the reagent). Slowly, aspirate and dispense this solution through the ZipTip for 3 minutes (**avoid air bubbles**). To quench unwanted sulfonation of OH-groups (Ser, Thr and Tyr) add 0.6 µl Stop- solution (50% hydroxylamin) to the CAF-solution, and aspirate and dispense through ZipTip about 10 times. Wash with 0.1% TFA (aspirate and dispense (to waste)10µl, repeate two times). Elute the modified peptide mixture by adding 3 µl 60% ACN/0.1% TFA in a

0.2 ml Eppendorf tube, aspirate and dispense carefully through ZipTip 5 - 10 times leaving the liquid in the tube.
This derivatization adds 136 Da to each peptide (if Lys side chain is blocked by O-methylisourea hydrogen sulfate (adding 42 Da) or, better, with 2-methoxy-4,5-dihydroxy-1H-imidazole (="Lys-TAG" from Agilent)adding 68 Da)

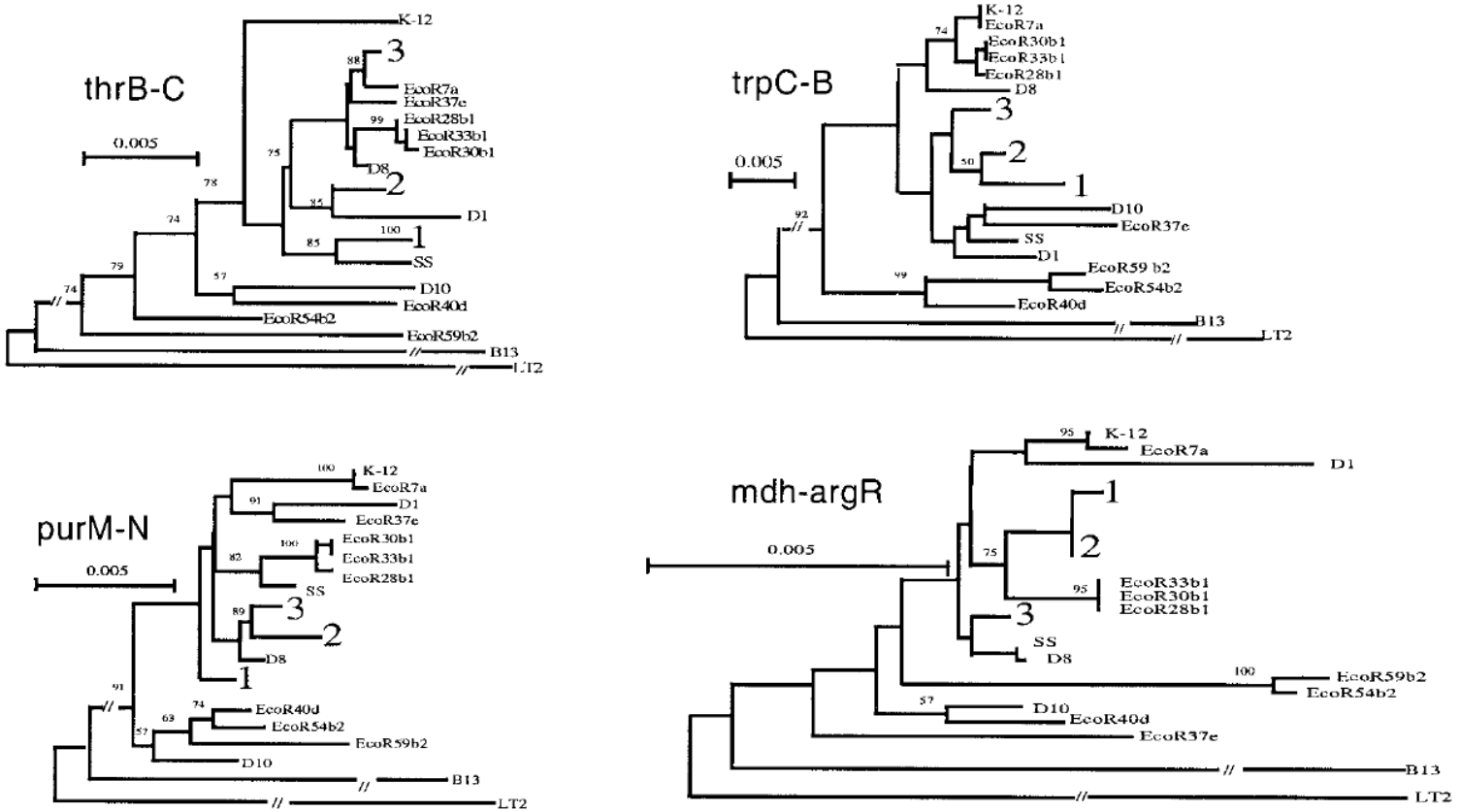
7. DD preparation of modified peptides and MALDI;

Mix 0.5-1µl of derivatizated peptides with 1µl DHB matrix (10-20 mg DHB/ml 60% ACN /0.1% TFA) and apply directly on a MTP 384 ground steel Target plate (BRUKER Daltonic GmbH, Germany). Apply peptide standard (see above) dissolved in matrix as nearest neighbor. (AnchorChip preparation (see point 3 above for details) works very nicely, and may be the preferable method if concentration of peptides is low).

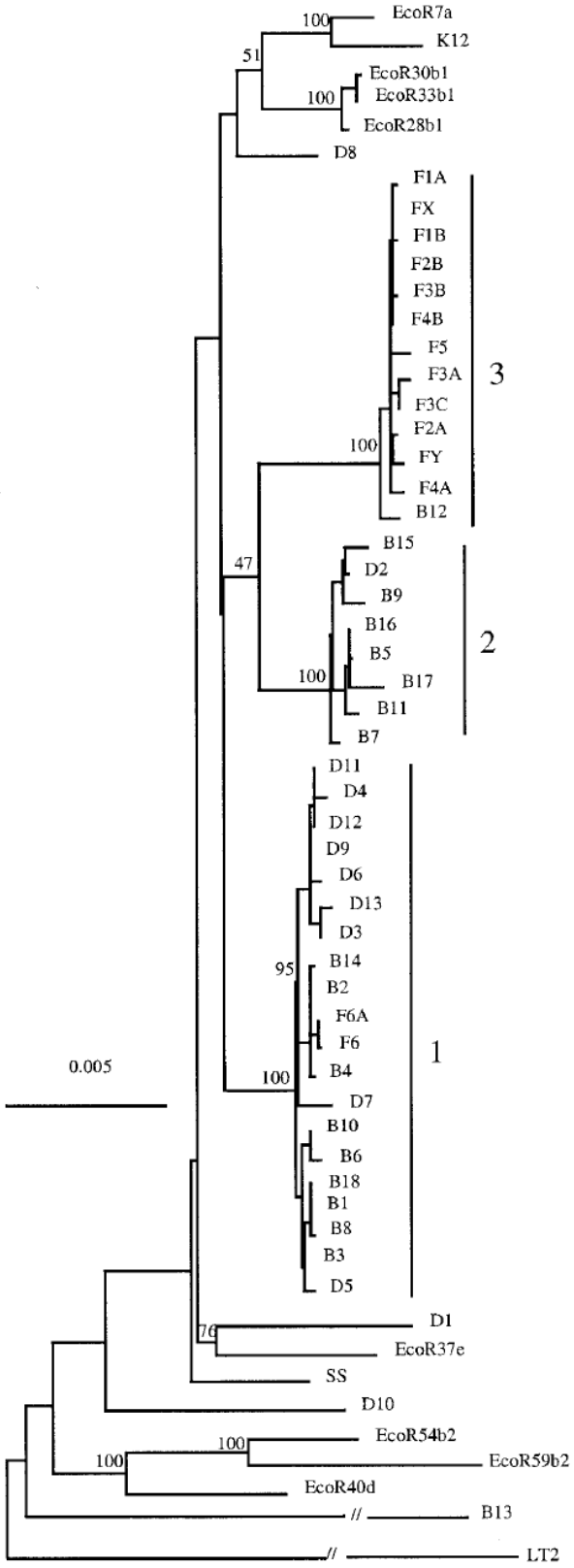
8. Analysis on MALDI ToF/ToF, DeNovo sequencing;

Produce a mass fingerprint of the modified sample, and calibrate (use external calibration as described above). Compare to the original spectrum. Look for peptide mass additions of 136 Da (C-terminal Arg-peptides, no Lys residues),
136 + 42 = 178 Da (C-terminal Lys-peptides, or C-terminal Arg-peptide with an internal Lys residue), and so on. Study the candidates, and select those that have a clean baseline near them - optimally ± 1% of precursor mass should be "free". Produce a PSD (Post Source Decay) spectrum of the precursor mass using the LIFT method offered by Ultraflex ToF/ToF (BRUKER Daltonic GmbH, Germany). The resulting spectrum should contain almost exclusively y-ions making *DeNovo* sequencing easy.

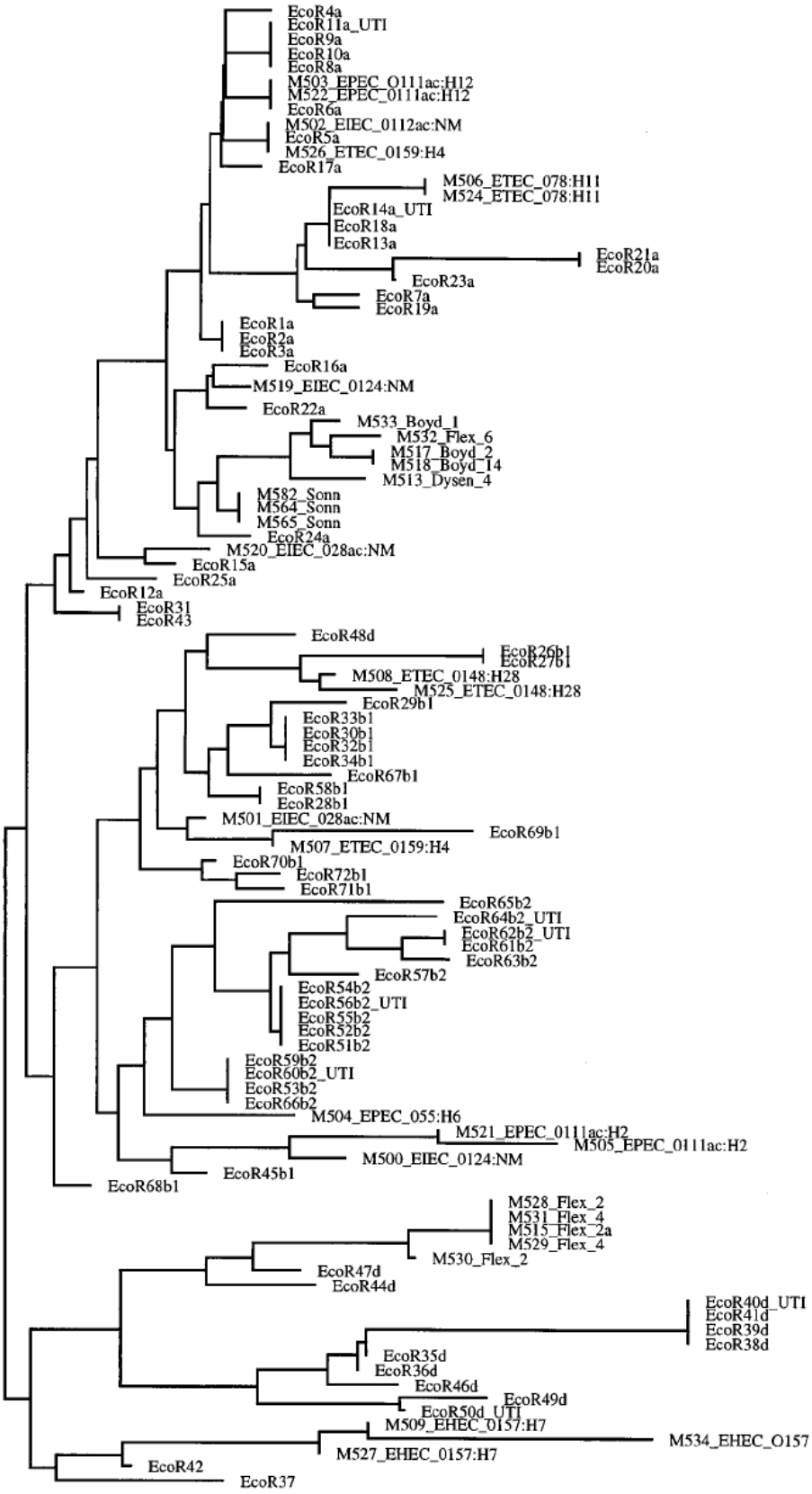
APPENDIX H



Phylogenetic trees made by neigbour joining method based on the sequences from four regions. Three major clusters of *Shigella* spp. are marked by number 1,2 and 3. No non- *Shigella* strains fell within any of the three clusters, but some representatives from the ECOR set of *E. coli* strains fell between major *Shigella* clusters indicating that the clusters has not arisen from one common, ancestral form (Pupo et al 2000).



Combined phylogenetic three (neighbor joining) of all the *Shigella* serotypes and eight *E.coli* from the ECOR collection based on 7160 bases from eight housekeeping genes in four regions.
Three clusters of strains were identified each including strains from more than one of the traditional *Shigella* speices. Five strains fall outside these clusters (SS, D1, D8, D10 and B13)
F: Flexneri, SS: sonnei, B: boydii, D: Dysenteriae followed by serotype number (Pupo et al. 2000)



Phylogenetic tree (neighbor joining) showing the genetic relationship of pathogenic E. coli strains of the commercial *E.* coli strains (ECOR set) resolved by MLEE.
The pathogenic groups (EPEC, ETEC, EIEC and UTI strains) of *E. coli* do not cluster except *Shigella* and the few groups of EHEC (Pupo et al. 1997)