Evaluation of a microarray-based approach for the detection of enteric viruses

Fadhila Al-tawqi

Submitted for the Degree of Master of Philosophy from the University of Surrey

Unis

Department of Virology
School of Biomedical and Molecular sciences
University of Surrey
Guildford, Surrey GU2 5XH, UK

2005

© Fadila
Chapter 1

1 Introduction

1.1 Enteric Viruses

Numerous enteric viruses replicate in the human gut. A considerable number of these viruses are now recognised as causes of foodborne and waterborne illnesses in human (Motarjemi et al., 1995).

Human enteric viruses can be categorised into groups according to the type of illness they produce. One group of viruses can cause gastroenteritis. The other group replicates in the human intestine and they migrate to other organs, causing hepatitis and meningitis, poliomyelitis, encephalitis. The viral families include Reoviridae, Caliciviridae, Astroviridae, Picornaviridae, hepatitis E virus and the Adenoviridae. With the exception of Adenoviridae, which have DNA genomes, all the other viruses have RNA. The reoviridae are double stranded RNA.

Rotaviruses, astroviruses, caliciviruses, and enteric adenoviruses have been reported as causes of viral gastroenteritis (Middleton et al., 1977; Kotloff et al., 1989; Esahli et al., 1991; Koopmans et al., 2000). Diarrhoea is largely due to the loss of cell lining in the intestinal tract although rotaviruses do induce an enterotoxin, protein NSp4. Enteric viruses are usually, but not always, shed in high concentration and levels in excess of $10^8$ per gram of diarrheal stool are often observed (Sobsey and Jaykus, 1991).

The impact of diarrheal disease on a worldwide scale is very significant. The burden of disease and death falls mainly on preschool-age children and largely those located in the developing world. Deaths when they occur are usually attributed to severe dehydration secondary to loss of fluids from diarrhoea and vomiting. However impaired absorption from the gut may continue for some time may exacerbate any other conditions and is especially serious if the child is already malnourished.

Transmission is generally considered to occur through the faecal/oral route, especially via high-risk foods such as shellfish or contaminated water. Infection via contact with contaminated surfaces can occur and occasionally a phase of airborne spread may
contaminate surfaces at some distance from the initial contaminating event, eg vomiting (Patterson et al., 1997; Marks et al., 2000). Water appears central to the transmission of many of these viruses, either leading to infection through ingestion or to the contamination of foods exposed to the water, eg shellfish through growth in the water, or irrigation/spraying or washing with contaminated water (Wyn-Jones & Sellwood, 2001; Seymour & Appleton, 2001; Carter, 2005).

The cost of these events to the community is likely to be very high. During a single outbreak in Colorado, when 5,000 persons were exposed to hepatitis A virus (HAV), the direct and indirect cost totalled half a million US$. However, no cost estimates were made for the implications in terms of working days lost and impaired performance, which would increase the cost of the outbreak to the community (Dalton et al., 1996).

Many enteric viruses are difficult to culture or are not yet culturable in the laboratory. Routine methods for detection of some enteric viruses are now established. Many are antibody based such as ELISA for astrovirus and more recently Norovirus and bead-agglutination for rotavirus and adenovirus. Although these tests are easily applied and commercially available, they are relatively insensitive when viruses are present in lower numbers ($<10^8$). In recent years, various PCR-based methods have been developed for detection of viral genetic material from clinical, food, and environmental samples. The latter assays are far more sensitive but they are very specific and these methods can detect only one or few pathogens simultaneously. Such an approach could be made more generally applicable if some means of distinguishing between amplification products could be provided. Microarray technology offers this possibility and has the potential to detect many pathogens simultaneously in the near future.
1.1.1 Properties of enteric viruses

Enteric viruses are diverse in size, ranging from the smallest at 22 nm to the largest at 110 nm. Their genomes, likewise, vary in size from 6.8 kb (astrovirus) to 38 kb (Adenovirus). The appearance and size of the main organisms is shown in figure 1.1. Since all viruses are obligate intracellular parasites, enteric viruses cannot replicate in food or in the environment. However, they can survive outside living cells for a limited period of time e.g. on inanimate surfaces, on hands and in dried faecal suspensions. These viruses are relatively resistant to heat, acidity of the stomach, disinfectants and low pH, approximately pH 3 (Green et al., 1998; Barry-Murphy et al., 2000; Bidawid et al., 2000, Sattar et al., 2000; Dolin et al., 1972). Viruses comprise nucleic acid genomes of either DNA or RNA surrounded by protein coats. In many viruses, especially those transmitted via the respiratory tract, there exists also an external lipid envelope. Although there are some enteric viruses with such a structure (e.g. transmissible gastroenteritis virus of pigs, a Coronavirus), most lack this envelope. This is probably because envelopes are fragile structures not suited to passage through and survival in the environment. However, few particles are needed to produce illnesses because transmission through the environment is always accompanied by dilution. Hence to compensate for their survival most enteric viruses infect at low dose. Infectious doses range from perhaps as low as 1 for Norovirus and less than 10 to around a 100 for rotavirus. (Moe et al., 1998; Koopmans & Duizer, 2004).

The genetic organisation of each virus group is different; here we will describe that of poliovirus as a basis for the discussion. Poliovirus is the best-characterised enteric virus and it provides a useful comparison for the consideration of other such viruses like the astroviruses, which will be described in more detail later. To begin the infectious cycle the viral particle must first encounter an appropriate receptor on the plasma membrane of a potential host cell.
Replication is illustrated in Figure 1.2. Poliovirus enters the cell by attaching to a host cellular receptor belonging to the immunoglobulin superfamily and named poliovirus receptor (PVR); an intercellular adhesion molecule, now identified as cell surface antigen CD155 (Daley, et al. 2005). Following attachment the virus is taken into the cell through the formation of a coated pit into which the capsid is engulfed and transported into the infected cell’s cytoplasm inside an endosome. This process also called receptor-mediated endocytosis is illustrated in (figure 1.2). The interior of the endosome becomes acidified triggering a change in conformation of capsid proteins, leading to externalisation and loss of VP4 and translocation of the virus RNA across the endosome membrane into the cytoplasm. The viral RNA is polyadenylated at its 3’ end and bears a covalently linked protein, VPg, at its 5 end. This protein is thought to serve as a primer for the synthesis of virus RNA and to remain attached to virus RNA that is encapsidated. For translation this molecule is removed. The virus RNA specifies one single open reading frame of viral replicase 3D which gives rise to all the virus proteins by a process of maturational cleavage (figure 1.2) All RNA viruses, except the retroviruses must provide an RNA-dependent RNA polymerase which is not present inside mammalian cells, in the case of polioviruses this is protein 3D which then catalyses to transcription of virus RNA. This occurs via the multicopy transcription of the incoming positive strand forming a structure termed replicative intermediate. This takes place on membranous bodies sequestered from the host cells endoplasmic reticulum and involves many other virus non-structural proteins (Vincent, 2001). These negative strands are then converted into more replicative
intermediates, this time synthesising new positive strands. Early in infection these strands are used primarily to increase virus translation but later on they become encapsidated into new virions that accumulate in the cytoplasm until viral protein induce cell lysis resulting in virus release.

In general terms, virus infected cells induce phenotypic changes referred to as cytopathic effects (CPE). In the case of poliovirus lysis occurs. Membrane integrity is lost and the cell may swell due to the absorption of extracellular fluid and finally, break open releasing more virions. It is important to realise that not all viruses induce this effect, although they may cause other CPEs.

For cell lysis to occur a phenomenon known as shutoff is exhibited. Shutoff is the sudden inhibition of host cell macromolecular synthesis. Host cell protein synthesis and RNA synthesis are inhibited by proteolytic digestion of the translation initiation factor eIF-4G so that ribosomes can no longer recognise capped mRNA, thus host cell become destroyed. Such modification leads to the translation of only uncapped poliovirus mRNA because its internal ribosome entry site (IRES) allows it to assemble the translation complex with the virus-modified ribosomes. This process serves two functions. First, to free up more ribosomes to translate the viral genomes, and second, to insure that the cell will break down and die, releasing the newly assembled viral particles.

In poliovirus-infected cells, this is the result of production of the virus 2A protein. This molecule is a protease, which cleaves the p220 component of eIF-4F, a complex of proteins required for cap-dependent translation of mRNAs by ribosomes. Since poliovirus RNA does not have a 5' methylated cap, but is modified by the addition of the VPg protein, virus RNA continues to be translated. In poliovirus-infected cells, the dissociation of mRNAs and polyribosomes from the cytoskeleton can be observed and this is the reason for the inability of the cell to translate its own messages. A few hours after translation ceases, lysis of the cell take place (Lamphear et al., 1995; Gradi et al., 1998).
Figure 1.2. The schematic representation of poliovirus replication cycle. Viral entry is by receptor-mediated endocytosis during which the virion proteins are sequentially removed, releasing virion-associated positive-sense RNA. This RNA is translated into a large polyprotein. Viral replicase released from the precursor protein then mediates generation of RI-1 and RI-2 to generate more mRNA that, unlike the original genomic RNA, has the VPg protein cleaved off. As infection proceeds, the replication complexes become associated with cellular membrane structures into replication compartments. Newly synthesized positive-sense RNA is also translated and the process repeats many times until sufficient capsid protein precursors are formed to allow assembly of the procapsid. Procapsids associate with newly synthesized positive-sense RNA still containing VPg at its 5' end, and entry of viral genomes results in capsid maturation. As the process continues, virions accumulate in the cytoplasm until viral proteins induce cell lysis and virus release occurs. The entire process can take place in the absence of a nucleus (Wagner & Hewlett, 2004)
Chapter 1

1.2 Human enteric viruses

1.2.1 Hepatitis viruses

There are two forms of enterically transmitted hepatitis: Hepatitis A and Hepatitis E (Cromeans et al, 2001).

1.2.1.1 Hepatitis A

Hepatitis A Virus (HAV) was classified, as enterovirus 72, but was subsequently removed as the only member of the genus hepatovirus in the Picornaviridae family. More recently a second virus Avian Encephalomyelitis virus (AEV) has been classified into the genus as well. In common with all picornaviruses HAV has a single stranded RNA genome of positive polarity that is approximately 7500 nucleotides long. Genome structure and replication scheme of HAV are similar to that of poliovirus and other member of the Picornaviridae family with the major exception that this virus shows greatly reduced cytopathology. It is less efficient at shutting down host cell protein synthesis and cell lysis is also reduced as the virus is secreted inside membranous vesicles rather than through rupture of the plasma membrane. These vesicles rapidly break up when the virus enter the bile and HAV particles thus released are non-enveloped of diameter ranging from 27 to 32 nm with 32 capsomeres on an icosahedral surface. It has a buoyant density of 1.33/1.34 g/ml and a sedimentation coefficient of 156 to 160 S (Coulepis et al., 1982)

Infection begins with ingestion of the virus via food or water. Particles pass through the intestinal lining with or without replication and reach the liver. Here they infect the hepatocytes. Replication is not particularly damaging to the cells, but the viral antigen evokes an immune response. Antibodies and cytotoxic T cells are produced which attack the infected cells in the liver with a view to clearing infection from the system. This immune attack induces the signs of hepatitis. These include fever, loss of appetite, nausea, and abdominal discomfort, often followed after a few days by jaundice. In general the severity of infection increases with the age of the patient at primary infection: infections in children under 5 years old tend to be mild or asymptomatic but those in adults are rarely so. However, asymptomatic infections or mild illness without jaundice is not impossible in adults.
Chapter 1

The illness begins 15–50 days or an average of 4 weeks after the virus is ingested (Mast and Alter, 1993). Symptoms may persist but normally resolve well. Some patients follow a relapsing course that may last 6 months or more but death is rare. Some patients experience permanent loss of some liver functions as in cirrhosis.

An example of the impact of HAV infection occurred in 1988 in Shanghai, China, leading to a very large epidemic (more than 300,000 cases), because adults ate partially cooked clams that were harvested from areas contaminated with raw sewage (Yao G, 1991; Halliday et al, 1991).

HAV infection occurs at greatly differing frequencies around the world and highly endemic areas. In the US, epidemics of hepatitis A have occurred with almost regular periodicity until the 1980s and are probably related to the age-specific proportion of the population that is susceptible to infection and the degree to which HAV is circulating in the general population or in the selected risk groups. The highest rates of HAV have consistently been in the western part of the country and are primarily in counties with a low population density and large Hispanic or Native American populations. In Native American populations 40% of children and over 80% of adults had evidence of infection (Shapiro et al., 1991; Shaw et al., 1990).

1.2.1.2 Hepatitis E virus

It has long been clear that HAV alone could not account for all the hepatitis of presumptive enteric origin and this gave rise to the concept of enterically transmitted non-A non-B hepatitis (Et NANB). After a long search Hepatitis E virus was identified in 1990, through molecular cloning (Reyes et al., 1990; Bradley, 1990; and 1992) and is now believed to account for practically all Et NANB. HEV is a small, nonenveloped virus that shares some morphological and biophysical characteristics with human calicivirus such as architecture and structure, and virion diameter ranging from 32 to 35 nm. The genome is a single stranded, positive sense RNA of approximately 7.2 kb.

Currently HEV is classified as a floating genus “The Hepatitis E like viruses” unassigned to any virus family (Berke and Matson, 2000). At least eight HEV genotypes have been described worldwide. Of these the human isolates of HEV can be divided into four genotypes having a divergence of not more than 20% of the nucleotides in open-reading frame (ORF) 2 and each with a distinct geographic distribution (Worm et al., 2002).
Infections in industrialized countries were found to belong to genotype (3, and 5 to 8), which are distinct from those described from the developing world (genotypes 1, 2 and 4). Worldwide, most HEV infections are caused by genotype 1, while the genotype 4 is recognised as the cause in China. Genotypes 1 and 2 have been associated with incidences in Burma and Mexico respectively.

Hepatitis E infection generally resembles that of HAV, with several significant exceptions. Firstly, the incubation period is slightly longer, with a range of 15 to 62 days. The longest incubation times are found to be up to 10 weeks (Mast and Alter, 1993). Secondly, the target age groups are generally young adults, rather than children with a particularly high fatality rate (up 27%) in pregnant women (Krawczynski et al., 2000; Worm et al., 2002). Lastly secondary attack rates are much lower than those of HAV.

The infection can occur in endemic, epidemic and sporadic forms and is often associated with contamination of water. It has been indicated that this disease may be endemic in developing countries, including countries in Asia, Africa, and Latin America. The endemcity is further suggested by the occurrence of 7- to 10-year epidemic cycles in central asia (Shakhgildian et al., 1986). Sporadically, cases in UK have been reported but are all imported by travelers to the UK from endemic areas, Asia, china, and Africa (Jothikumar et al., 1993). Interestingly some 2% of the Western population may have antibody to HEV- like viruses (Meng et al., 2002) despite this lack of obvious disease. HEV-like viruses have also been detected in sewage draining areas in which there is no clinical HEV (Meng et al., 1997). This apparent paradox may be solved by the detection of HEV like viruses in supermarket pig liver (Yazaki et al., 2003). Urban homeless have the highest rate of seroprevalence (Smith et al., 2002). HEV is also known to be present in boar and deer liver and consumption of raw or poorly cooked foods have led to human infections (Matsuda et al., 2003). However pigs’ liver and its pattern of consumption in the West is presumably an inefficient vehicle for infection. If and when it does occur, it seems to be subclinical (Yazaki et al., 2003).
1.2.2 Viruses causing gastroenteritis

1.2.2.1 Enteric adenovirus

Human adenoviruses are members of the *Adenoviridae* family and occur in 51 distinct serotypes, classified into six different subgroups, A-F, on the basis of immunological, biochemical and biological differences. Within those groups, subgroup F, serotypes 40 and 41 have been found to be regularly associated with gastroenteritis in infants and young children. Types 42 and above have only been reported in association with HIV. These viruses tend to cause diarrhea, which is less acute than that caused by rotavirus but is of longer duration (Wadell, 2000).

Adenoviruses are 80 nm in diameter, non-enveloped, and contain double-stranded DNA of approximately 35000 bp. They are stable at low pH, and unaffected by bile and gut enzymes, allowing them to replicate to high titers in the gastrointestinal tract.

Clinically, adenoviruses, (mainly those of subgroups B, C), cause respiratory tract infections (tonsillitis, laryngitis, bronchitis and pneumonia). In the past enteric adenovirus (subgroup F) infections were thought to be second most common cause of infantile diarrhoea after rotaviruses (Krajden et al., 1990). Of the adenoviruses detected in faeces, 30-80% are types 40 and 41(Krajden et al., 1990). In large surveys, adenovirus infections were associated with diarrhoea in children in approximately 3% of all cases tested (Waters et al., 2000). Adenovirus infections with gastroenteritis are a frequent event after bone marrow transplantation (Hale et al., 1999a).

Enteric adenoviruses cause diarrhoea that lasts 6 to 9 days and may be associated with fever and vomiting. Usually, adenovirus gastroenteritis is mild but severe colitis has been found in immunocompromised patients (Krajden et al., 1990). Many enteric adenovirus infections are asymptomatic (Van et al., 1992).

Enteric adenoviruses types 40 and 41 are widespread and cause endemic diarrhea and outbreaks of diarrhea in hospitals, and childcare centers (Chiba et al., 1983; Van. et al., 1992). Enteric adenoviruses are reported to cause 2% to 3.8% of pediatric diarrhea, in inpatients and outpatients, mainly in children below the age of two years (Brandt et al., 1985, De Wit et al., 2001). Adenoviruses are frequently isolated from sewage, seawater
and shellfish (Girones et al., 1995; Pina et al. 1998), but food-associated adenovirus outbreaks are apparently unknown, maybe because of childhood immunity that continues throughout life rather than the absence of these viruses in food.

1.2.2.2 Enteroviruses

Enteroviruses are a genus of the family *Picornaviridae*. They have been shown to be present in human faeces and in domestic sewage (Cliver 1994). There are 68 immunologically distinct serotypes of the human enteroviruses including the polioviruses serotype 1-3, Coxsackievirus of group A viruses (24 serotypes), coxsackieviruses B (6 serotypes), echoviruses (30 serotypes), and enterovirus 68-71 with no specific subgrouping (Melnick, 1996; Muir et al, 1998). Enteroviruses consist of an almost featureless capsid of 27 nm in diameter containing 60 protomers, each possessing 4 surface proteins, VP1, VP2, VP3, and VP4 arranged with icosahedral symmetry to enclose a genome of a single stranded RNA of positive polarity of 7-8.5 kb in length.

Enterovirus virions are stable at pH 3 and remain viable at −70°C for years, and at 4°C for days to weeks. The viruses are also resistant to ethanol and lipid solvents, but they are inactivated at temperatures of 55-60°C for several minutes. Enteroviruses can also be inactivated by ultraviolet light (UV), formaldehyde, hydroxylamine (Kocwa-Halich, 2001). Shedding of viral particles in the faeces can occur as long as 3 months post-infection (Block & Schwartzbrod, 1989).

These viruses can be transmitted through the oral-faecal route and many have the ability to target the central nervous system (CNS). They have been shown to cause infections such as poliomyelitis, meningitis, and encephalitis. They have also been associated with diseases such as Diabetes Mellitus, myocarditis as well as, acute hemorrhagic conjunctivitis. In the case of polioivirus, vaccination has almost eradicated the virus in developed countries. However, it remains a problem in developing countries. There are no vaccines for other enteroviruses. Peaks for the condition affecting the CNS occur in summer and early autumn in temperate climate and all age groups can be infected. Seasonal patterns are not seen in the tropics (Maguire et al, 1999). In general, most of the members of the enterovirus group could be transmitted via food. Although all may spread only infrequently by this route, only a small number of foodborne disease outbreaks caused by Coxsackie and echoviruses have been recorded (Cliver, 1997).
1.2.2.3 Rotavirus

Rotaviruses, a genus of the *Reoviridae* family, contain a genome composed of 11 segments of double stranded RNA (dsRNA) enclosed in a triple layered protein. They are approximately 70 nm in diameter, non-enveloped with a smooth outer edge. The virus consists of a core containing the RNA genome and inner and outer capsid exhibiting icosahedral symmetry (Pesavento et al., 2003).

Rotaviruses are classified into seven distinct groups, A–G. But only groups A, B and C have been identified in humans. Two outer capsid proteins, VP7 and VP4 elicit neutralizing antigens and define the virus G (glycoproteins) and P (protease sensitive) serotypes, respectively. Currently, at least 21 P and 14 G serotypes have been identified in humans (Esters, 2001).

Group A rotaviruses have been identified as the major etiological agents of severe infantile diarrhoea worldwide, and cause approximately 2 million hospitalizations and 352000–592000 deaths in children <5 years of age each year (Parashar et al., 2003). Infection generally confers long-term immunity to serious gastroenteritis but asymptomatic or minimally symptomatic re-infection can occur throughout adult life (Bishop, 1996). Gastroenteritis has been observed in elderly residents of nursing homes with several fatalities reported, but these viruses do not appear to be a major cause of morbidity in adults (Lew et al., 1991; Gangarosa et al., 1992; Kapikian, 1996; Koopmans, 1997). Immunocompromised patients are also at risk. Older children and adults are resistant to rotavirus infection, probably due to active immunity, reinforced by repeated infection throughout life.

In contrast to group A rotavirus, infection with group B and C viruses has caused outbreaks of disease in older children and adults. Group B rotaviruses have caused epidemics in China affecting all ages (Fang et al., 1989). Group C rotaviruses occasionally cause outbreaks of infection in children. Rotavirus infection is seasonal in temperate climates with peaks in the cooler months (Cook et al., 1990). Group A rotavirus have frequently been detected in shellfish and water (Le Guyader et al., 1994; Gilgen et al., 1997). However, there has been no documented disease following shellfish consumption (Lees 2000).
1.2.2.4 Caliciviruses

Caliciviruses are members of the family Caliciviridae. They are enteric, nonenveloped viruses with icosahedral symmetry. Capsids are built primarily from a single protein and show cup-like depressions on their surface formed by rings of arch-like protrusions (Prasad et al., 1999). These cup shaped depressions accumulate stain and appear dark under electron microscope (EM). It is these that give the virus its name (calici “a cup”). The genome consists of a linear, positive sense single-stranded RNA molecule of 7.4–8.3 kb (Clarke and Lambden, 1997; Green et al., 2001). The Caliciviridae family has been divided into four genera, and only two of them infect humans. These are Norovirus (NoV) (formerly “Norwalk-like viruses”, NLV) and Sapovirus (SaV) (formerly “Sapporo-like viruses”, SLV) (Green et al., 2001; Mayo, 2002).

1.2.2.4.1 Noroviruses

Noroviruses comprise a single molecule of positive-strand RNA contained within a capsid built from a single type of capsid protein (termed VP1). Minor amounts of a second protein (VP2) are also present. The diameter is about 34 nm and the surface often appears fuzzy under EM. Cup-like depressions on the surface are rarely clear (Carter et al., 1991).

NoVs may infect all age groups but primarily include children and adults (Kaplan et al., 1982). Nowadays NoVs are recognized as the most common cause of nonbacterial gastroenteritis outbreaks in these age groups (Koopmans et al., 2000). Clinical features of NoV mediated gastroenteritis include an incubation period of 12-48 hours (mean 36 hours), which is relatively short for a virus. Symptoms include nausea, vomiting, severe abdominal cramps, and diarrhoea. Data from the CDC indicates that vomiting is relatively more prevalent among children, and a greater proportion of adults experience diarrhoea (CDC, 2001). The illness is usually over in 12-60 hours, but faecal shedding of the virus sometimes continues for more than a week. During illness, intact viral particles may be found both in the diarrheal stool and in the vomitus. Frequently, the amount of virus shed is under the limit of detection by direct EM, and immune EM may be needed to identify the viral particles. Infection leads to production of antibody against the virus, but immunity apparently does not last much more than a year (Jaykus et al., 1994). Some studies have indicated that people with antibody against a NoV were more susceptible than those without antibody (Johnson et al., 1990; Jaykus et al., 1994). Recently, Hutson
and his group have investigated the relationship between a person's blood type and the risk of NoV infection and symptomatic disease after clinical challenge. It was found that individuals with an O phenotype were at significantly increased risk of NoV infection whereas individuals with a B histo-blood group antigen were resistant to NoV infection. They believe the presence or absence of receptors or other genetic factors may explain such resistance or susceptibility, which could be useful in developing antiviral treatments for the elderly or immuno-compromised patients. In conclusion, the study suggested that some people are more susceptible genetically to NoV infection, which might help explain the fact that certain members of the population experience NoV infection and disease after re-exposure to the virus even when displaying high antibody levels to NoV (Hutson et al., 2002).

1.2.2.4.2 Sapovirus

Sapoviruses are genetically distinct from the NoVs (Liu et al., 1995; Green et al., 2000) and may include several distinct strains (Jiang et al., 1997). SaVs differ from NoVs by causing symptomatic infections predominantly in infants and young children (Cubitt, 1989; Roman et al., 2002), although outbreaks of gastroenteritis in older people have been reported (Noel et al., 1997). Seroprevalence studies in both developed and developing countries indicate a high prevalence of antibody in young children (>80%), with immunity appearing to be long-lived (Lambden et al., 1994, Parker et al., 1994, Lew et al., 1994). Clinical symptoms are similar to those following NoV infections when they occur in persons of the same age. In younger individuals SaV tends to induce milder disease than in older persons, and diarrhoea may predominate over vomiting (Lee, 2000). Infection is common in institutional settings such as schools, day care centres and nursing homes.

It is not clear whether SaVs are of significance in foodborne infections and more studies are needed to clarify their role. Genotypic analysis is being used increasingly to investigate the NoVs and SaVs respectively (Maguire et al., 1999; Koopmans et al., 2000; Buesa et al., 2002; Zintz et al., 2005).
1.2.2.5 Astroviruses

Astroviruses are members of the *Astroviridae* family. They are 28nm non-enveloped viruses possessing a single stranded RNA of positive polarity as their genome. Virus particles are small, at 35-40 nm diameter, and round. They display a characteristic five or six pointed star-shape in their centres (Kurtz and Lee, 1987). There are eight recognized human serotypes. Human astroviruses cause sporadic individual cases and occasional outbreaks of diarrhoea illness mainly in infants, young children and the elderly. Astrovirus infections occur throughout the world and in temperate zones are most common during the winter months. Astroviruses have been detected by EIA in the stools of 1.5% to 9% of children with diarrhea and in 1% to 2% of stools from asymptomatic control children (Glass et al., 1996). Outbreaks of diarrhea associated with astroviral infection have been reported in schools, pediatric hospitals, and child care centers (Oishi et al., 1994, Mitchell et al., 1993, 1999).

The study of astrovirus serotypes in various populations has demonstrated that type 1 was the prevalent type in the United Kingdom (UK), accounting for more than 65% of cases (Noel et al., 1994). More recently survey indicated that other types are increasing in the UK; this change in types may be associated with increase in the number of outbreaks (Willcocks et al., 1995). Multiple types co-circulate in the USA and Mexico (Noel et al., 1995, Guerrero et al., 1998; Mendez et al., 2000). In Japan, large outbreaks of food-borne astrovirus type 6 involving more than 1000 children and adults have been reported. This shows the need to evaluate differences in geographic and temporal distribution of types (Oishi et al., 1994).

Clinical symptoms are milder than those following NoV infection, but do include vomiting, diarrhoea, fever and abdominal pain. The incubation period may be longer than for NoV at 2-4 days with illness lasting another 2-3 days but occasionally 10-14 days (Kurtz and Lee, 1987).

Antibodies to astrovirus are acquired by 75% of children by 5 to 10 years of age (Kurtz and Lee, 1978). Myint et al 1994 demonstrated that 93% of surfers had serologic evidence of exposure to astrovirus antigenic type 4, whereas a control group of non-recreational users had only 22% seroprevalence to the same type. This report implicated water exposure as a risk factor for infection by astroviruses in surfers.
Astroviruses have been implicated in outbreaks where food, water, and seafood (oyster) were vehicles of infection (Kurtz and Lee, 1987; Oishi et al., 1994; Yamashita et al., 1991; Anon, 1998; Le Guyader et al., 2000).

Since astroviruses will form the main subject of this thesis their genetic organization is described in more detail below.

The genome of astroviruses is approximately 6.8 kb in length, excluding the poly A tail at the 3' end, and encodes three open reading frames (ORFs), called ORF1a, ORF1b and ORF2 (Willcocks et al., 1994) (Figure 1.3). ORF1a and ORF1b at the 5' end encode non-structural proteins (nsP) including a serine protease and a polymerase, respectively. The ORF2 at the 3' end encodes the capsid precursor protein (Lewis et al., 1994). ORF1a is approximately 2,842 nucleotides long and overlaps with ORF1b by 70 nucleotides. It is 1542bp long. This area is highly conserved among human astrovirus serotypes. The virus produces both a full-length genomic and a subgenomic RNA transcripts. Proteins from ORF1a and 1b are all made from the full-length transcript; there is (-1) ribosomal frameshifting event between the two frames. Structural proteins are made from the subgenomic RNA that contains only ORF2 (Monroe et al., 1991; 1993; Willcocks et al., 1993; Lewis et al., 1994). The length of ORF2 varies according to strain, but is between 2,358 and 2,388 nucleotide long.
Figure 1.3 Genome organisation of astrovirus.
The ORF1a and ORF1bs: Non-structural regions encoding the serine-protease (SP) and RNA-dependent RNA polymerase (Pol): ORF 2 structural region encoding the capsid proteins.
1.3 Transmission of the human enteric viruses

Enteric viruses are spread through the faecal oral route, and some have been shown to be present in food, sewage effluents (Sair et al., 2002). Four major routes of transmission have been recognized. Transmission can be by food (e.g. shellfish), and water, or by person-to-person transmission (e.g. infected food handlers) or vomitus.

Most of the enteric pathogens can be transmitted by person-to-person spread, particularly in the very young and the elderly, those suffering from learning difficulties, and in circumstances where normal hygiene measures are difficult to maintain such as overcrowding or military manoeuvres. The extent to which viruses are transmitted by this route varies with the virus; transmission of rotavirus and adenoviruses has only rarely been associated with food and these agents are thought to spread almost entirely by the person-to-person route. In contrast Norovirus transmission in the community was attributed almost entirely from person-to-person spread rather than to infected food and water entering the home (FSA 2000).

Drinking water is another route by which humans become exposed to waterborne pathogens, as these viruses are not necessarily destroyed by water preparation for drinking especially if aggregated. Reports for the sensitivity of these viruses to chlorination show that most non-aggregated viruses would be destroyed. One report of a remarkable resistance of norovirus to chlorination (Keswick et al., 1985) is now thought to be in error (Doultree et al., 1999). Similarly these viruses survive wastewater treatment and contaminate receiving waters. Here they may spread to humans through recreational use of water or indeed if this water is used to irrigate or spray crops or even if they are fertilised with inadequately decontaminated sewage sludge (Metcalf et al, 1995). Such concerns apply mainly to produce eaten without peeling or cooking and this includes mainly soft fruits and salad vegetables. A prime source of virus infection is the consumption of shellfish, (oysters, clams, mussels, cockles), grown in waters that have been contaminated with sewage effluents either through inadequate treatment or temporary overwhelming of treatment plants through heavy rainfalls. These organisms concentrate the viruses within their bodies through their filter feeding activity and as they are often consumed raw or with minimal cooking they present an obvious risk of infection.
Thirdly, infected food handlers who practice poor personal hygiene are an extremely common source of contamination. For example, foodborne disease surveillance data from Minnesota from 1984 to 1991 indicated that viral gastroenteritis was the most common foodborne illness, accounting for 39% of the reported outbreaks, predominantly linked to infected food handlers (Hedberg and Osterholm, 1993). Most of these workers could have been ill either prior to or at the time of food preparation or they could be asymptomatic.

Most foods involved in outbreaks for which human handling is a factor are the ready-to-eat (RTE) foods. RTEs are foods that are supposedly fit to be eaten without washing, cooking, or additional preparation by the consumer or by the food establishment. The contamination of RTE and prepared foods comes from poor hand-washing practices of food handlers after toilet use, as faecal material can be left on hands or even under nails, which then can come into contact with food products (Jaykus 2000). Handling cooked products with bare hands is a factor responsible for pathogen transfer to RTE foods (Bryan, 1995).

Fourthly, transmission can occur through exposure to vomitus. In this case, attack rates can be related to the distance between secondary cases and the primary case that vomits (Marks et al., 2000). This is consistent with airborne spread of the virus and infection by inhalation and subsequent ingestion of virus particles. The vomitus-oral route of transmission is a likely explanation of how NoVs are spread so rapidly and why propagated epidemics frequently take place.
1.4 Detection methods

The development of Detection methods for enteric viral infections has been driven largely by the requirements for clinical diagnosis. As such they have been developed for application to clinical specimens in which virus concentration is usually high. These methods are not generally applicable to food, water or environmental samples. Food vehicles were identified only by common exposure studies following an outbreak. Where such methods lacked specificity they were also of little use in implicating the same virus in related cases and this led to difficulties in linking illnesses if patients had become ill from a virus originating from a common source. For this reason these methods were not suitable to pre infection screening of foods and so they were of little use in ensuring food safety or to indicate the extent of foodborne illness outside of outbreak scenarios where clear evidence of common exposure or even of the existence of a problem was difficult to obtain. This gap has been bridged with the development of nucleic acid amplification techniques such as Polymerase Chain Reaction (PCR) and NASBA. However these new and greatly more sensitive methods were generally only applicable to a search for a few viruses at a time. Nevertheless, the advent of molecular technology specially the polymerase chain reaction (PCR) has revolutionized the detection of enteric viruses from clinical, food, water or environmental samples. It has become the gold standard.

PCR relies for distinction on the specific amplification of sections of nucleic acid from defined pathogens. Observation of amplification is then taken as indicative of the presence of that pathogen. The amplicon thus formed can be further investigated usually by sequencing or hybridisation with a detector probe to identify the virus concerned and in this way to assess the relatedness of viruses that may have caused geographically well-separated outbreaks. This is a relatively laborious process and was never likely to be readily applicable to the prospective screening of food, water or environmental samples for the dozens of human viruses that might be present, and hundreds that could be there if animal infecting strains were to be included.

Recently, an advance in diagnostic technology has emerged through the arrival of microarray technology. This technology permits the resolution between many different viruses based on hybridisation to a probe array in which hundreds or perhaps thousands of virus sequences may be represented. This means that a total virus population might be
amplified either by culture or PCR or both. The identification of viruses contained in the mixture can then proceed simultaneously by hybridisation of the mixture of amplification products to a screening array spotted with the relevant probes.

The objective of this project is to investigate the suitability of such an approach using the human and animal astroviruses as a pilot study. The astroviruses are closely related in their non-structural genes and thus longer probes tend to yield similar base match values with all viruses. Thus a secondary objective was to see if positive hybridisation could be obtained using probes small enough to show more specificity for some types of virus. In order to appreciate the advantages of this approach the conventional virus identification is described briefly below.

1.4.1 Cell culture

The traditional method for detection of viruses involves cell culture. It is expensive, labour-intensive and time-consuming. Culture has never been used routinely in the diagnosis of enteric virus infections. There are no methods for the laboratory culture of some of the most important enteric viruses such as Noroviruses and Sapoviruses. Other viruses such as hepatitis A virus (HAV) can be grown but are slow growing and difficult to isolate. Rotaviruses can be cultivated but of the three types (A, B and C) that infect humans, only group A are cultivable. For all of these reasons diagnosis has traditionally used electron microscopy. Indeed, for many (rotaviruses, caliciviruses and astroviruses) it is electron microscopy that led to their original discovery.

1.4.2 Electron microscope (EM)

This is a “catch-all” method and no prior knowledge of the virus is necessary to detect it. Historically it is thought that this method has identified adenovirus and rotavirus infections quite well but has been less successful when applied to smaller viruses that are less distinct or variable in appearance or that may be present in lower numbers. EM is relatively insensitive because it requires virus concentrations of at least $10^6$ per ml (Atmar et al., 2001). Direct electron microscopy requires skilled microscopists in such cases but has nevertheless been used to screen specimens for enteric viruses in the public health laboratories of many countries (Svensson et al., 1983, Monroe et al., 1991, Maguire et al., 1999).
Electron microscopy has been extended to incorporate antibodies for the identification and classification of viruses. This technique is called immune electron microscopy (IEM). This uses antibodies to clump viruses of indistinct morphology thus enabling their detection against a background of general debris. Furthermore since the reaction with antibody is dependent on protein structure this technique was type specific and was used to provide serotype information even for non-cultivable agents. Solid Phase Immuno Electron Microscopy (SPIEM) has now superseded IEM. Gough and Shukla (1980) first described it, whereby solid grids were coated with antibody as means of capturing the virus on a solid phase. SPIEM has been used to detect and classify Norovirus (Lewis et al., 1995; Hale et al., 1999b). Such modifications have been made to improve the ease with which viral particles are detected and to simplify the performance of the test. However the lack of sensitivity means that EM has never been a suitable technique for searching viruses present in low amounts as in food or water and its usefulness has been limited to faecal samples. As our knowledge of the viruses concerned has improved, and the supply of reagents becomes more routine, electron microscopy is being gradually replaced with ELISA-based methods and PCR (Atmar et al., 2001).

1.4.3 Immunoassays

The principles behind these methods depend on the use of an antibody to capture the virus from a sample mixture whilst other components of the sample can be washed away. The two approaches are briefly described below.

1.4.3.1 Radioimmunoassay (RIA)

Radioimmunoassay (RIA) was developed as an alternative to IEM. It has been used to detect NoV antigen in stool (Greenberg & Kapikian, 1978), and HAV in contaminated drinking water (Hejkal et al., 1982). However, owing to the use of radioisotope, the method is not popular and has largely fallen into a state of disuse. It will not be discussed further.

1.4.3.2 EIA/Enzyme Linked Immunosorbent Assay (ELISA)

EIA has been developed with the advantage that it does not require radioactive materials. It is a versatile and widely used method that can be applied to the detection of antigens. When the antigen-antibody reaction takes place on a solid surface, the reaction is also
known as enzyme-linked immunosorbent assay (ELISA). In this assay a “capture” antibody specific for the viral antigen being sought is bound to a reaction surface, such as the wells of the microtitre plate or the surface of a plastic bead. When the specimen is added, viral antigen present in the specimen binds to the capture antibody. After washing to remove unbound specimens material, a second antiviral antibody, the detector antibody, is added. The detector antibody can carry an enzyme label (figure 1.4A) or can be detected by the addition of a third antibody (figure 1.4B).

Commercial ELISA-based assays are available for detection of group A rotaviruses, adenoviruses and astroviruses, and for some lineages of NoV in clinical samples (Vipond et al., 2000; Kobayashi et al., 2000). However these methods lack sensitivity for direct application to environmental or foods samples and also suffer from problems imposed by the antibodies available. These tests are difficult to apply generally unless there are pan-reactive epitopes, which enable a wide variety of viruses to react with relatively few sera or antibodies. In the case of many (such as the Noroviruses for instance) the immune response elicited tends to be narrow and thus any sera raised tend to have a restricted reactivity. Pan reactive epitopes are known for astrovirus (Herrmann et al., 1990) and there is some evidence that broad if not universal epitopes exist for Norovirus (Jiang et al., 1995), but in the case of the Noroviruses these form a relatively small component of a polyclonal response and most antibody generated is type specific. The situation is continuously improving, but even for those not yet readily detectable by these means (non-group A rotaviruses, HAV, SaV and the remaining NoV), the diagnosis can be made by detection of viral nucleic acid by molecular assays (Gouvea et al., 1991; De Leon et al., 1992; Vinjé et al., 2000) and to this end polymerase chain reaction (PCR) methods have been established.
Figure 1.4 Enzyme immunoassay (EIA) for antigen detection. A, Direct; B, indirect. E, Enzyme.
1.4.4 Nucleic Acid Amplification

1.4.4.1 Polymerase Chain Reaction (PCR)

PCR is a method for the \textit{in vitro} synthesis of specific DNA sequences. Most polymerases require a short region of double stranded nucleic acid for initiation of synthesis. The method utilizes a pair of synthetic oligonucleotides as primers, each binding to the 5' of one strand and to the 3' end of a double stranded DNA (dsDNA) target, which will be exponentially reproduced. The hybridized primer acts as a substrate for a DNA polymerase, which creates a complimentary strand via sequential addition of deoxynucleotides. The PCR is a cyclical process as depicted in figure 1.5. The PCR reaction is capable of amplifying extremely low quantities of nucleic acid. Thus, it increases sensitivity whilst maintaining specificity to a particular viral target.

However, the majority of the enteric viruses are single stranded-RNA agents. Therefore, they can be detected by first copying the target RNA to its complementary DNA (cDNA). This is accomplished using the enzyme reverse transcriptase (RT), hence the name RT-PCR. It is a rapid and highly sensitive method that has been extensively developed for the qualitative and quantitative detection of RNA viruses. This approach has suffered from the laborious pre and post PCR handling steps required to evaluate the amplicon (Guatelli et al., 1989).

Post PCR success depends on analysis that proves that the amplicon detected is in reality the expected target sequence. This requirement for post PCR analysis has often limited such tests to a search for only a few viruses at a time. Nevertheless, PCR technology can be used in a number of ways to identify and characterize viral nucleic acids. This approach is not perfect and certainly doesn’t detect all viruses, but it offers sensitivity far in excess of that provided by other methods and in the absence of successful culture procedures offers the only approach for the detection of viruses in very dilute samples such as food, drinking water or the environment.
1. Denaturation of the template DNA
94°C

2. Annealing of the Oligonucleotide primers
50-60°C

3. Synthesis of new DNA
74°C

4. Repeat the cycle 25-30 times

Figure 1.5 Schematic diagram showing the basic steps in Polymerase Chain Reaction
1.4.4.2 Alternative Nucleic Acid Amplification procedures

RT-PCR assays were developed to detect viruses in clinical samples or foods targeting individual viruses only. An improvement has been the multiplex approach, which attempts to overcome the limitation of conventional PCR. In this format, several different primers or primer sets are included in a single reaction mix each targeting a different virus type. The primers are so chosen as to generate a specific size of amplicon. Gel analysis or probe technology can then identify the amplified viruses.

Multiplex RT-PCR offers several advantages over the monoplex methods, such as reduced labour, reagent costs and faster detection (Chamberlain et al., 1994). A multiplex RT-PCR has been reported for the simultaneous detection of Norovirus, Astrovirus and Adenovirus in clinical samples (Rohayem, et al., 2004). Unfortunately, the use of several different primer sets may adversely affect the efficiency of the nucleic acid amplification as the reaction effectively sets up a competition between all possible amplification reactions. Each reaction may compete with the others at varying efficiencies under differing conditions and according to the composition of the virus mixture being analysed. There are also interactions between the primers, especially as the number of sets of primers in the reaction increases, thus leading to a decrease in the overall sensitivity of the assay. In addition, it is limited to the detection of no more than six viruses.

Alternative nucleic acid amplification methods are also available. Of particular interest may be Nucleic Acid Sequence Based Amplification (NASBA) (Sooknanan, 1995). This method uses three enzymes (reverse transcriptase, RNase H, and T7 RNA polymerase) and two specific oligonucleotide primers (one of which bears T7 promoter sequence added to its 5’ end to amplify target RNA or DNA sequences (Compton, 1991).

The method has advantages over PCR because it is more rapid, can be made quantitative, and the amplification is carried out at a single temperature, avoiding the need for specialized thermocycling equipment. However, it is also limited to the detection of no more than six viruses at time.
1.4.4.3 Confirmatory Methods

Traditionally, the presence of an amplicon has always been confirmed by agarose gel electrophoresis and identified as the specific target largely on the basis of size. However, to eliminate the problems of false positives and non-specific amplifications, alternative confirmation methods can be applied. Such confirmation methods include nested PCR, restriction endonuclease digestion, and RT-PCR amplicon sequencing but these methods can take up to 48 hours to complete and may be subject to cross-contamination.

Within the last few years, techniques based on linking ELISA approaches to DNA hybridization have resulted in the potential of faster and quantitative confirmation. In these methods, a solid phase capture is linked to immunomolecular detection using chromogenic, chemiluminescent, or fluorescent substrates (Andreolelti et al., 1996). More recently, fluorescent DNA probes such as molecular beacons (Tyagi et al., 1996) and the Taq-Man assay (Le Cann et al., 2004) have been reported for the real-time measurement of hybridization kinetics. Taq-Man seems to be an ideal choice except for the fact that it cannot perform more than three assays in one reaction. A lot of effort has been placed on reducing overall PCR confirmation time by bypassing gel electrophoresis. This was achieved by focussing specifically on the investigation of the use of fluorescent dye-labelled probes and enzyme detection strategies. The ultimate goal was to achieve endpoint detection within a shorter period.

The critical analyses of the advantages and disadvantages of the methods discussed above have led to the design of a microarray system to detect, type and confirm amplicons of a larger number of viruses simultaneously. It can also be useful in the discovery of viruses. Therefore, to develop and evaluate a microarray system may compensate those limitations towards providing a cheaper and rapid detection method.
1.4.5 Microarray

Microarray technology is based on the principle of hybridising labelled nucleic acid in solution in the sample (targets) to probes immobilised on a solid support. In essence this is very similar to conventional dot-blot technology using nitrocellulose or nylon membranes, the difference lying in the scale and size of the dot array that can be produced and the accuracy with which it can be read. This is achieved using complementary sequences or oligonucleotide (probes), immobilised on a solid surface, (usually glass or silica). This presented a considerable technological challenge for reproducible application to complex mixtures of probes representing several hundred viruses in a sample. This improvement was driven by the development of technology applicable for whole genome screening in the postgenomic era. Membrane based arrays may be in the format of line probe blots. Scanning or imaging the microarray surface reveals the results of hybridisation. This will be dependent on the fluorescence intensity of the sample compared to the control.

There are two major approaches in the construction of DNA microarrays. The first method is the physical attachment of DNA fragments such as library clones or polymerase chain reaction (PCR) products to a solid surface (Cheung et al., 1999). These sequences are referred to as probes or features. Each spotted DNA array can have a density of up to 5000 features/cm². The features comprise dsDNA molecules (genomic clones or cDNAs) up to 400 nucleotides in length and must be denatured prior to hybridisation. This microchip is too laborious and expensive for common practice.

The second approach uses single-stranded synthetic oligonucleotides manufactured in situ on the plate using photolithographic techniques (Lipshutz et al., 1999). These arrays can have a density of up to one million features/cm²; each feature comprising up to $10^9$ molecules of a single stranded oligonucleotide varying from 25-70 nucleotides in length. Owing to the reduced hybridisation specificity of a short oligonucleotide, 20 features can represent each gene on an array. The advantages of oligonucleotide microarrays are that they are designed based on sequence information alone and there is no need of intermediate such clones, PCR products or cDNA. The process of making an array for relevant applications is illustrated in figure 1.6.
The processes of sample preparation are broadly similar for any application. After extraction of RNA populations, the RNA must be labelled with fluorescent dyes (figure 1.6). This is commonly achieved through direct synthesis of cDNA by reverse transcriptase, or preparation of amplified DNA products by RT PCR. In either case label can be incorporated by the use of specifically labelled primers that add label only at the ends of the molecules synthesised, or through the incorporation of labelled precursors during enzymatic synthesis, which will introduce label throughout the target molecule. A third possibility is to chemically label the target after it has been transcribed in vitro although this is more usually applied to RNA targets. The labelled samples are then dispensed onto the microarray for hybridisation.

Nevertheless, each sample preparative procedure has its benefits and limitations. In an ideal situation, assuming time is not a critical factor, the post PCR labelling approach can be used because of the larger amount of DNA available as opposed to first-strand cDNA labelling which is quicker. However, this approach requires much more starting RNA.

1.4.5.1 Microarray Applications

The major application of microarray technology is in screening tissues for changes in their levels of mRNA expression under differing environmental conditions. This demands that two samples be produced, a control prepared under standard conditions and a test sample prepared under the conditions of interest. Both samples are processed and labelled with different fluorescent dye; the two commonly used are Cy3 and Cy5. This mixture is then hybridised to the array and the resulting colour intensities binding to each spot are measured by laser scanning. This allows the user to see whether the intensity of each colour has increased or decreased relative to the other under the conditions of test. As a detection system only one sample is needed, prepared from the sample under test. Negative controls are clearly desirable but there is often no perfect control available and careful consideration of exactly what might constitute an adequate control will be necessary to ensure that specific hybridisation is occurring.
Figure 1.6 An illustration of the processes involved in making and using an array. The top represents the preparation of probes from the viruses of interest to the spotted arrays. The bottom represents the DNA or RNA extraction from samples or infected cell line, amplified and labelled with either Cy3-dCTP (green) or Cy5-dCTP (red). When applied to the array bearing the immobilised probes, the target binds to complementary sequences. An example of an array result is shown on the right: the green spots represent hybridisation of the probe only with target sequences labelled with Cy3-dCTP; the red spots represent hybridisation of the probe only with target sequences labelled with Cy5-dCTP; the yellow spots represent hybridisation with both target sequences. Adapted from (Clewley 2004)
DNA microarray technology is highly promising for diagnosing viral infection as evaluated by Shih and colleagues (Shih, et al. 2003). Two oligonucleotide probes were made from the gene of the enterovirus 71 and were spotted onto glass slides. RNA extracted was amplified by RT-PCR. The amplification reaction mixtures contained the two sets of primers and the RT-PCR products were hybridized together to the probes in the array. Of 71 enteroviruses tested 67 were found to yield a positive signal showing that this novel method has sensitivity of about 90% and its specificity was 91%. In addition, other benefits of this system included a relatively short length of the amplified fragment, which makes it appropriate for immediate hybridization without any fragmentation. The hybridization procedure took 2 hours and the whole procedure was completed in 6 hours. It also minimizes the risk of contamination.

Since microarray analysis has been shown to be promising in diagnosis, other researchers have extended this principle to include molecular typing as well. The design and construction of rotavirus microarray was based on sets of capture oligonucleotides of about 21 bp in length and specific for the G and P types of rotavirus, which are immobilized on glass microscope slides (Lovmar et al., 2003). With these oligonucleotides as probes, the RT-PCR products containing type-specific regions of the VP7 and VP4 genes are captured on the microarray by hybridisation under low-stringency conditions. The same oligonucleotides with 3’ ends that are matched to the sequences of the captured templates were extended with a mixture of dNTPs containing Cy5 labelled dUTP using a thermostable DNA polymerase. The fluorescence incorporated in the sequence-specific primer extension reaction was measured by the microarray scanner. The method detected human rotavirus in 40 samples tested and correctly determined both the G and the P genotypes of 35 of the 40 strains analysed. The results of genotyping fully agreed with the results obtained by sequencing and sequence-specific multiplex RT-PCR. Thus showing the flexibility of microarray technology.

Despite the fact that multiplex RT PCR is limited to target only few pathogens, about 6 viruses, Boriskin and coworkers illustrated a possible microarray procedure for the detection of a range of viruses in central nervous system (Boriskin et al., 2004). The microarray consisted of 38 gene targets for 13 viral causes of meningitis and encephalitis. The PCR products were made from these gene targets and were spotted in quadruplicate onto glass slides. All primers were 22-mers with similar melting temperatures and G+C
contents averaging 48%. A mixture of the viral products was used as a control. Multiplex PCR- products from the DNA and RNA viruses in cerebrospinal fluid (CSF) and non-CSF specimens were labelled with a fluorescent dye, Cy3-dCTP. When tested with fluorescently labelled DNA, the expected correct homologous hybridisation were observed on the array as defined by the mean hybridization signal intensity compared to the background signal intensity. Of 60 clinical specimens processed by multiplex PCR-microarray procedure 41 gave pattern for a specific virus on the array (23 echoviruses, 4 herpes simplex virus [HSV-2], 4 varicella-zoster [VZV], 1 human herpes virus [HHV-7], 1 HHV-6, 2 HHV-6B, 3 Epstein-Barr virus [EBV], 1 polyomavirus [JC], and 2 cytomegalovirus [CMV]). However, this microarray consisted of spotted arrays of PCR products, which is too laborious and expensive for common practice. Therefore, oligonucleotide array-based would be the method of choice.

A further advantage of this procedure is that it may assist in the analysis of uncharacterised viruses. This is done by scraping hybridized viral sequences from a DNA microarray spot, followed by amplification, cloning and subsequent sequencing (Wang et al., 2003). So far, microarray strategies have been applied to clinical samples only. The increasing amount of new sequences in GeneBank database has helped to design a selection of arrays for viruses.

Since a water quality microarray has not yet been developed for the detection of viruses in water, the ultimate goal would be to design a multiplex PCR microarray for the detection of viruses in environmental samples and with the possibility of incorporating probes for all enteric viruses.

This project will evaluate this approach using the human and animal astroviruses. Since these are closely related, small probes need to be used to detect regions of sufficient similarity but sufficient difference for distinction. This is a compromise between sensitivity and specificity and relatively short oligonucleotides to develop a pilot test array. This will then be evaluated for detection of human viruses in cultivated material and if time permits it would also be tested using environmental samples.
1.5 The objective of the study

The project has the following specific objectives:

1. Design probes capable of distinguishing between human and animal astroviruses and if possible between serotypes of human astrovirus.
2. Cultivate human astroviruses in vitro for the preparation of control RNA material for labelling using cell culture and in vitro transcriptions from cloned cDNA
3. To consider different methods for preparation of labelled samples
4. To seek regions of sequence similarity for the pan-specific PCR amplification of human and animal viruses from environmental mixtures.
Chapter 2

2 Materials and methods

2.1 Sequence Analysis

Sequences of virus genomes of interest were obtained from Genbank. They were then subjected to Clustal-X alignment in order to identify regions suitable for the design of probes for microarray use and primers for PCR. In this design the primers were intended to be broadly cross-reactive and thus achieve amplification from as many potential viruses as possible, whilst the probes were designed to be more specific and to generate signals from a restricted number of viruses to aid identification. This was not always possible. In order to maximise sequence conservation between viruses we targeted the RNA dependent RNA polymerase gene of the viruses. This is well conserved within a virus family, has amino acid sequence motifs essential for its function, which can be used to design pan-reactive primers. Furthermore although this protein is conserved in some regions, differences do exist elsewhere. These can be found because the high level of conservation means that this is a frequently targeted gene and thus many sequences are available in the database allowing us to identify these regions of difference.

2.1.1 Design of PCR primers and oligonucleotide probes

Sequences of the RNA-dependent-RNA polymerase gene (Pol) of human astrovirus (Accession No NC_001943, L23509, AF292074, AF292075, AF292076, AF292077, AF248738, AF260508) and animal astroviruses (AY179509, NC_002469) were obtained from NCBI Genbank and aligned with the ClustalX program (Thompson, et al. 1994). Prior to alignment the sequences were saved in the FASTA format. Regions that were conserved between all human serotypes, and animal astroviruses were identified by examining the alignment similarity scores and were then used to design probe for microarray and primers for RT-PCR amplification. They were tested for specificity by using them as BLAST query sequences against the sequences in GenBank. This
confirmed that high-scoring sequences belonged only to the astroviruses. Specific oligonucleotide probes were selected using the following criteria: length approximately 40nt, GC content 40-50%.

PCR primers were then synthesised by Genosys (London Rd., Pampisford Cambs, UK). These were diluted to 226 µM on arrival and used in the PCR reactions to a final concentration of 0.226µM (section 2.5.1). Probe oligonucleotides were custom made and purchased from Operon biotechnologies (Nattermannallee 1, 50829 Cologne Germany) and then immobilized on a modified glass surface to create the microarray.
2.1.2 Microarray printing and processing

Oligonucleotide probes were synthesized commercially (Operon biotechnologies) and were suspended in 6x print buffer at concentration of 20 mM. They were spotted in quadruplicate on Gamma Amino Propyl Silane (GAPS™) coated slides using a robotic arrayer (QArray, Genetix). After spotting, the probes were cross-linked to the slide with ultraviolet irradiation, baked at 80°C for 2h, and stored in closed containers until application. The arrangement of spots on the array is shown in figure 2.1. The printing was kindly performed by Vassilios Mersinias (University of Surrey).

Fig 2.1 Arrangement of spots on the microarray slide. RED: Controls, Green: DMSO controls, Dark blue: HAstV probes, Light blue: Animal astrovirus, Orange: HAV, Sea green: BEV, Black: Empty spots.
2.2 Cells and Virus

2.2.1 Cell culture

CaCo-2 cells were maintained in minimal essential medium (MEM) invitrogen, Gibco BRL, England) (Eagle’s modification with Earles salts). Medium was supplemented with 2mM L-glutamine, 2%(v/v) non-essential amino acids, penicillin (100 U/ml) and streptomycin (100 µg/ml) (Gibco) and finally supplemented with 10% foetal calf serum (FCS) for cell growth. Cells were propagated in 25 or 80 cm² tissue culture flasks at 37°C in an atmosphere of 5% CO₂. Cells were passaged by removing the medium in the flask, washing the cells with trypsin/versene to remove medium and serum that would inhibit the trypsin and stripping the adherent cells from the monolayer by the addition of a small quantity of trypsin versene (0.025% EDTA in PBS). Cells sheets were incubated at 37°C for 10 min to detach and once they had done so they were released from the bottle surface with a sharp tap on the bench before they were resuspended by gentle pipetting, diluted with MEM and aliquoted into fresh flasks. Typically bottles were split on a 1-to-4 ratio.

2.2.2 Virus strains

Viruses used in this study are human astrovirus types 1 to 6. Type 1 HAstV A2/88 Newcastle was isolated and supplied by Dr M M Willcocks. All other serotypes were reference strains isolated at John Radcliff Hospital, Oxford and supplied by Professor J B Kurtz, Oxford UK. When required cloned cDNA copies of virus RNA were also provided prepared from the strains above and supplied by Dr M M Willcocks (University of Surrey).

2.3 Target preparation for cDNA labelling

2.3.1 Propagation of astrovirus in cell-culture

Human astrovirus was propagated in vitro in CaCo-2 cells maintained as above. The receptor for HAstV in these cells is believed to lie on the basolateral cell membranes and becomes inaccessible once the cells reach confluence (Willcocks et al., 1990). Thus infections were carried out at approximately 80% confluence judged by eye. Monolayers were washed twice with serum-free MEM prior to infection. The astrovirus isolate
(0.5ml) was then inoculated onto the cell monolayer and supplemented with serum-free MEM (0.5ml) containing 5μg/ml of trypsin. The cells were incubated at 37°C with gentle agitation for 1 hour, followed by the addition of 4.5 ml serum-free MEM, 100μl 50x trypsin, and monitored daily for cytopathic effect (CPE). Infected cells were harvested at 48 hours by three cycles of freeze/thaw to disrupt the cells to maximise release of the virus particles. The medium was then clarified at 2500 rpm (559 x g) for 10 min, aliquoted, snap frozen and stored at -80°C. Initially one vial of virus was used to produce a master stock. From this a sub-master could be produced and finally this was used to grow working stocks. When required fresh working stocks could be prepared by growth of the appropriate sub master stock as necessary.

2.3.2 RNA grade materials

Care was taken to reduce the risk of contamination with RNAses. Sterile, disposable plasticware was used wherever possible. Any glassware used was pre-treated with RNaseZAP (Sigma). All solutions were prepared with RNAse-free glassware and DEPC-treated sterile distilled water.

2.3.3 RNA extraction

Cells were harvested for RNA preparation when CPE was visible in infected cells (or at appropriate times for control cells). Briefly, the cells were harvested by scraping into the medium, and pelleted by centrifuging at 2500 rpm (559 x g) for 10 min. The medium was removed and the cell pellet resuspended in 6ml cold RNAzol B (AMS biotechnology, UK). Chloroform (600μl) was added and the mixture was shaken vigorously for 15 sec, placed on ice for 5 min and then centrifuged at 14,000 rpm (8765 x g) for 15 min. The upper (aqueous) phase was transferred to a fresh tube and an equal volume of isopropanol was added. The mixture was stored at 4°C for 15 min and centrifuged at 12,000 rpm (6440 x g) for 15 min to pellet the total RNA. The supernatant containing DNA and proteins was discarded and 75% (v/v) ethanol was added to wash the pellet. After vortexing the mixture was further centrifuged at 14,000 rpm (8765 x g) for 5 min. The supernatant was discarded and the pellet was dried at 37°C for a few minutes. The pellet was dissolved in 200μl of water, and reprecipitated by addition of 2.5 volumes of absolute ethanol and 1/30th volume of 3M of sodium acetate (pH 5.0).
The mixture was chilled at -20°C overnight. RNA was collected by centrifugation, resuspended in RNase-free water, and used for reverse transcription RT-PCR, labeling, and hybridized to a microarray.

### 2.3.4 Determination of RNA Concentration

The extracted RNA concentration was determined spectrophotometrically by the nanodrop, ND 1000 (Agilent technologies) or a classical spectrophotometer (UV 1101, Biotech photometer). The procedures followed were as described in the manufacturers’ protocol. An OD of 1.0 at 260 nm in a 1cm silica cell corresponds to a concentration of 40 ug/ml for RNA.

### 2.3.5 Preparing of labelled cDNA from total RNA

2 μl oligo-dT (20 μg/μl Amersham) were added to 9 μl of extracted total RNA sample. The mixture was placed at 70°C for 5 min, and cooled on ice for 10 min to allow the primer to anneal. For the synthesis of first strand cDNA, a mixture comprising 4 μl 5x CyScript buffer (Amersham), 2μl DTT (0.1 M stock, Amersham Biosciences), 1μl optimized nucleotide mix (Amersham), 1μl dCTP CyDye-labelled nucleotide, and 1μl CyScript reverse transcriptase (100 U/μl) was prepared. This was added to the primer annealing reaction mixture and incubated at 42°C for 1.5 hours. The labeled cDNA was either stored at -20°C or used for immediate purification. Prior to purification, 2 μl of 2.5 M NaOH was added into the labeling cDNA and incubated at 37°C for 15 min to remove excess RNA. Finally, adding 10μl of 2M HEPES neutralized the reaction.

### 2.3.6 Pre-hybridization

Microarray slides were placed in a pre-soaking and pre-hybridisation solutions using the Pronto Universal Microarray Hybridization Kit (Corning, USA) as described by the supplier’s instructions. It is efficient at eliminating background fluorescence (Raghavachari, et al. 2003).
2.3.7 Hybridization of labeled cDNA and analysis

Labeled cDNA samples (Cy3) were mixed with 16 µl of hybridization solution (Corning, USA) and denatured at 95°C for 5 mins, centrifuged at 14,000 rpm for 2 mins. The labeled cDNA samples were then added to the surface of the microarray, which had been immersed in the prehybridisation solution. The slide was covered with a glass coverslip pre-washed with nuclease-free water, followed by ethanol. The slide with the coverslip was placed in a sealed hybridization chamber. 15 µl of 3x SSC was added to every well at the two ends of the chamber, which was then placed in waterbath. After hybridization at 42°C for 16 h, the slide was taken out and washed as described for the Pronto! Universal Hybridization kit (Corning, UK). The array was then dried by centrifugation at 2,500 rpm ((559 x g) for 2 mins. Scanning was performed using a 428™ Array Scanner (Affymetrix) at 550nm.
2.4 Target preparation for RNA labelling

2.4.1 Restriction digestion reaction

The cloned cDNA provided by Dr M M Willcocks was digested using specific enzyme. The reaction was performed in a 50μl reaction volume containing 5 μl 10x enzyme buffer (Promega), 5 μl 10x BSA, 20ng/5 μl recombinant plasmid DNA, 36 μl sterile distilled water and 2.5 μl of PstI enzyme (1u/μl Promega), and the reactions were incubated at 37°C for 2 hours for the old vector (pTZ19), and the new vector pGEM-3Z (Promega) which contain two promoters. In the latter case, 1μl calf intestinal phosphatase enzyme (1u/μl) (Boehringer Mannheim) was added and incubated at 37°C, and 56°C for 15 min respectively. The restriction fragments were separated on a 1.2% agarose gel. Both the insert and vector were excised, and extracted or purified separately using the QIAquick Gel Extraction kit (Qiagen).

2.4.2 Ligation reaction

The purified insert was ligated into a plasmid vector pGEM-3Z (Promega). The ligation reaction contained in final volume of 20 μl: 9 μl of the insert DNA (4ng/μl), 2μl of 10x ligation buffer (Promega), 3 μl linearized vector (35 ng/ μl), 3 μl distilled water, 2 μl polyethylene glycol 6000 (25% PEG), and 1 μl T4 DNA ligase (3u/μl). The reaction was mixed by pipetting and incubated overnight at 16°C.

2.4.3 Transformation Reaction

The E.coli DH5α competent cells were prepared by Marian and kept in the liquid nitrogen (university of surrey). One vial was taken from the liquid nitrogen and placed on ice to thaw. The LB broth was allowed to reach room temperature. A water bath was equilibrated to 42°C and LB plates warmed to 37°C. The tube containing the ligation reactions were briefly microfuged and 7μl of the reaction was added to a vial of the 50 μl E.coli DH5α competent cells and mixed by stirring gently with the pipette tip. The vial was incubated on ice for 30 minutes, then heat shocked in a 42°C waterbath for exactly 45 sec, and subsequently placed on ice for 5 min. 200 μl of preheated (at 42°C) LB Medium
(Appendix) were then added to each tube, incubated at 37°C for 1 hour to allow expression of newly acquired antibiotic resistance genes and then placed on ice. The transformation mixture was divided into four equal aliquots and spread onto LB agar plates containing 250 μl IPTG, 100 μl x-Gal and 200 μl ampicillin (Appendix). The plates were allowed to dry for 10 minutes at room temperature, then inverted and placed in a 37°C incubator overnight. The plates were examined the next day for the growth of white colonies.

2.4.4 Plasmid isolation using Miniprep kit

White colonies were picked and transferred into universals containing 5 ml of LB medium (Appendix). The cultures were incubated overnight at 37°C with vigorous shaking and centrifuged at 13000 g for 2 minutes to pellet the cells. Plasmid DNA was extracted from 4.5ml of overnight culture using a miniprep kit as described in the manufacturer’s protocol (Qiagen, UK).

2.4.5 Screening plasmids DNA by restriction digestion

To determine whether the extracted plasmids DNA contain the insert of the correct size, they were screened by restriction enzyme digestion using PstI enzymes to excise the inserted DNA. The orientation of the insert was confirmed using EcoRI of enzymes.

Two reaction was set up in 20 μl volume containing 10x enzyme buffer (Promega), 2 μl 10x BSA (10mg/ml, Promega), 2 μl recombinant plasmid DNA, 13 μl sterile distilled water and 1 μl of PstI (1u/μl), EcoRI enzyme (12u/μl Promega) respectively. The reactions were mixed gently by pipetting, and centrifuged for 30 seconds at 13000 g. The reactions were incubated at 37°C for 2 hours. Agarose gel analysis was performed using a 1kb DNA marker.

2.4.6 In vitro transcription of RNA

From the purified plasmid DNA, 8μg of plasmid DNA was Linearised with the appropriate restriction enzymes required to cleave the DNA distal from the promoter site to be used. In general after NheI, BamHI, or SacI endonucleases (10u/μl, Promega) were found suitable and permitted the cleavage of the vector on opposite sides of the inserted
This allowed us to prepare run-off transcripts from either SP6 or T7 promoters, which flank the insert in order to generate transcripts of uniform size and opposite polarity. In this work NheI was used for positive sense RNA transcription from the T7 promoter and BamHI or SacI for negative strand transcription from the SP6 promoter. A sample of each Plasmid template was analysed on a 0.8% Agarose gel to confirm that linearization was complete before transcription. RNA transcription reactions were then carried out using the MEGAscript kit (Ambion, UK) as described in the manufacturer's protocol. The RNA product was then purified using Rneasy MinElute cleanup kit (Qiagen). The concentration of the RNA product was determined using nanodrop, ND 1000 (Agilent technologies).

2.4.7 Preparation of labelled RNA

0.88 µl of the transcribed sense RNA (1µg), 1 µl 10x labeling buffer (20 Mm, Amersham), 1µl Cy3 and, and 7.12 µl RNase free water were added to make a total volume of 10 µl. Reactions were gently mixed and incubated at 37°C for 1.5 hours. The labeled RNA was then purified using GFX PCR DNA and Gel Band purification kit (Amersham). The incorporation of Cy3 into RNA was quantified by the nanodrop ND1000 (Agilent technologies). Similar reaction was performed for antisense RNA with Cy5. Cy3 and Cy5-labelled mRNAs were dried down using a speedVac™ separately. They were then dissolved in 6 µl nuclease free water. The two samples were denatured at 95°C for 3 min, and cooled on ice for 30 seconds and were ready for hybridization.

2.4.8 Hybridization of labeled RNA and analysis

6 µl of each labeled RNA samples were mixed with 7.5 µl of microarray hybridization buffer (Amersham, Biosciences UK) and 15 µl of 100% (v/v) Formamide, and centrifuged for 30 seconds. The mixture was then added to the surface of the microarray slide. The slide was placed in a sealed hybridization chamber; 15 µl of 3x SSC was added to every well at the two ends of the chamber. After hybridization at 42°C for 18 h, the slide was taken out and washed with 1x SSC, 0.2% (w/v) SDS for 10 minutes, then twice with 0.1x SSC, 0.2% (w/v) SDS for 10 minutes. Finally, it was dipped into distilled water for 10 seconds and gently dried with air stream. The slide was scanned using a 428 Array scanner (Affymetrix) at 550nm.
2.5 Target preparation for DNA labeling

The aim of this procedure was to randomly amplify the DNA of any given sample, which would, as much as possible, be representative of the sample.

2.5.1 RT PCR amplification

The isolated RNA from section 2.5.1 was amplified using the same primer sets designed. The strand cDNA was synthesized by incubating 5 μl of the extracted total RNA with 1.6 μg/ml of antisense primer in a total volume of 50 μl reaction containing 200 units of reverse transcriptase (Promega, USA), 40 units of RNAsin (Promega, USA), and 5 mM dNTPs at 37°C for 60 min.

An aliquot (10 μl) of the first strand product was used as template for amplification in a 100 μl reaction containing, 2.5 μM of sense primer, 2 mM of dNTPs, and 1 unit of Taq DNA polymerase (Promega, USA). Amplification, was performed in a Techgene DNA thermal cycler, and involved an initial denaturing step at 94°C for 1 min, followed by annealing at 55°C for 30 seconds and extension at 72°C for 30 seconds. After completion of the cycles, a final extension at 72°C for 10 min to fill any “ragged ends” was carried out. PCR products were electrophoresised in a 1.2% agarose gel in TBE buffer and visualized by ethidium bromide staining under ultraviolet irradiation.

2.5.2 Agarose gel electrophoresis

After amplification, the DNA products were analysed by electrophoresis in an agarose gel. Initially, a stock solution consisting of 50 ml 10x TBE (Appendix) was diluted 1/10 in MQ water. The gel was prepared by mixing 0.6 g of Agarose (Sigma, molecular grade) to 50 ml from prepared solution. The mixture was then carefully microwaved for 2-3 minutes to boil dissolving the agarose powder, and allowed to cool to approximately 65°C. 2.5 μl of Ethidium Bromide (0.5 μg/ml) was added before the gel was poured. Meanwhile, 5 μl of gel loading buffer (20% glycerol, 0.1% SDS, and a small spatula-full of bromophenol blue), and 4 μl of MQ water were added to 1 μl of each PCR product. DNA size markers were prepared by adding 5 μl of a 100 base pair DNA ladder (Promega, U.K.) to 1 μl of gel loading dye. The markers and the PCR products were then loaded onto the gel. The gel was run at 100 volts, for 45 minutes. Ethidium bromide stained DNA bands were then visualized by fluorescence under ultraviolet light.
2.5.3 Purification of PCR product

PCR products were purified from Agarose gels using the QIAquick gel extraction kit (QiAgen) as described in the manufacture’s protocol. The DNA product was accurately measured by nanodrop, ND 1000 (Agilent technologies) as described in the manufactures’ protocol.

2.5.4 Preparation of labelled DNA

Enzymatic digestion of genomic DNA can sometimes produce better results. The pure DNA was denatured and then labeled using a mixture of Klenow fragment and Cy3. The final volume contained 2 µg of DNA (10µl), 1 µl random primers, and 41.5 µl distilled water. The reaction was incubated for 5 minutes at 95°C, and then cooled on ice. Then, 5 µl 10X Klenow buffer (Promega.), 1 µl dNTPs (5Mm each), 1.5 µl Cy3-dCTP, and 1µl klenow fragment (5u/µl), mixed and incubated at 37°C for 5 hours. The labeled DNA was then purified using GFX PCR DNA and Gel Band purification kit (Amersham). The sample was dried in a SpeedVac at room temperature and redissolved in hybridisation solution. The incorporation of Cy3 and Cy5 into DNA can then be quantified with nanodrop. Finally, the sample was hybridized on the array as in section 2.3.6.
Chapter 3

3 Results

3.1 Design of PCR primers and oligonucleotide probes

The initial stage in evaluating the microarray was to design primer sets and probes specific to the virus of interest. It was decided to develop sets for human and animal astroviruses separately because of the lack of conserved regions between them. Human astroviruses were chosen as models since they are cultivable in mammalian cell lines.

The probes for the microarray were synthesized from human and animal astroviruses. For the design of the probes and primers, a region of the viral genome was selected on the basis of two criteria:

(1) To find high conserved region between all eight human astrovirus serotypes so that a clear hybridization signal could be expected.

(2) The existence of conserved areas surrounding the probe regions to facilitate the design of primers that could be used for multiplex amplification targeting all human astrovirus serotypes in one sample (e.g. environmental samples).

Considering the first criteria, the RNA-dependent-RNA polymerase gene was selected and two Human astrovirus and one animal astrovirus specific probes were designed. Each had a length of 40bp and GC content varying from 32.5%-45%. Regarding the second criteria, one degenerated forward and reverse primer sequences were designed to target all the eight serotypes for the first and second strand syntheses. Another primer sets to target Mink and Ovine astrovirus (3.1. and 3.1.2). The resulting primer set was reactive to HAstV-1, 2, and cDNA as expected (see figure 3.6).
### Chapter 3

#### Primers

<table>
<thead>
<tr>
<th>RdRp region</th>
<th>Sequence 5'-3'</th>
<th>Tm</th>
<th>Position*</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAstV-Fwd</td>
<td>GGT GAA GTC ACC TTG CAG ACG CG</td>
<td>74°C</td>
<td>3787-3809</td>
</tr>
<tr>
<td>HAstV-Rev</td>
<td>GTG GTC CTC TGC CAT GAA CGC AG</td>
<td>74°C</td>
<td>4222-4245</td>
</tr>
<tr>
<td>Mink-Fwd</td>
<td>GAGGAGTTTCAGACAGATGC</td>
<td>60°C</td>
<td>1-23</td>
</tr>
<tr>
<td>Mink-Rev</td>
<td>CTTCCACGACTCCGTAG</td>
<td>66°C</td>
<td>220-240</td>
</tr>
</tbody>
</table>

3.1 *Nucleotide positions relative to reference sequences. The primer set was designed for the detection of the eight human astrovirus serotypes and Mink and Ovine astrovirus. The Tm was within the required criteria.

#### Capture Probes

<table>
<thead>
<tr>
<th>RdRp region</th>
<th>Sequences</th>
<th>Position*</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAstV-1</td>
<td>TGGTACAGGCTTTGAGTTTGCTTATTTTAATGGACCAG</td>
<td>3865-3904</td>
</tr>
<tr>
<td>HAstV-1</td>
<td>AACCTTGACCGCTGTGCAAGCCTTTATTTTAATGGACCAG</td>
<td>4105-4144</td>
</tr>
<tr>
<td>Mink &amp; Ovine</td>
<td>CAGGCATTTGAGTATGCCTATATCCCATCCAGAGAAGACG</td>
<td>172-213</td>
</tr>
</tbody>
</table>

3.1.2 *Nucleotide positions relative to reference sequences. The astrovirus probes immobilized on a microarray.

#### 3.2 Microarray design

The microarray was prepared by spotting synthesized oligonucleotides of the conserved sequences of human and animal astroviruses onto the surfaces of glass slides by using a Qarray (Genetix). Controls were immobilized at the same time. The control system was composed of three parts: (i) Empty controls were DMSO-only and lacked gene fragments, (ii) negative controls were Lucidea Universal ScoreCard controls, which were not homologous with astrovirus, and (iii) Positive controls were the astrovirus. Eight Blocks of the 8 x 8 microarrays could be simultaneously immobilized on a glass slide. The pattern of printing is shown in figure 2.1.
3.3 Propagation of astrovirus in cell culture

A human astrovirus-isolate was successfully propagated in a continuous line of human colonic carcinoma (CaCo-2) cells line. Infection was indicated by the presence of cytopathic effect (CPE) in the monolayer of infected cells compared to a control flask. CPE was seen in the infected cultures after two days. The cells on the edges of the sheet were the first to develop CPE and become rounded and clustered. This effect then spread throughout the culture and progressed to sloughing of the monolayer. When CPE was extensive the cells were harvested and extracted.

3.4 Analysis of microarray hybridized with labelled cDNA

After CPE was detected visually, the cells were harvested for the extraction of RNA, which was measured spectrophotometrically. The concentration of RNA obtained (2.64 μg/μl), which was just about sufficient (2.5-25 μg) for first-strand cDNA labeling. The Cy3 labelled cDNA did not produce any hybridization with the probes.

For positive hybridisation the colour would change from red to green as measured spectrophotometrically at 550 nm. However, no significant colour was observed at all i.e. a black background lacking any coloured spots was visible.

Owing to the lack of positive results from the first-strand cDNA labelling, an alternative approach was adopted. It was in vitro transcription of RNA and its direct labeling after transcription. This necessitated the subcloning of the astrovirus sequences into a transcription competent vector.

3.5 Cloning of ORF1b and in vitro transcription

This technique allowed the synthesis of cloned RNA. Unfortunately, the pTZ19 cannot be used because it does not contain two promoter sites that are required to produce the sense and antisense RNA. The vector selected was pGEM-3Z because it has two promoters, T7 and SP6, for the transcription of both the sense and antisense RNA.

The pTZ19 vector containing the HAstV-ORF1b was digested by the enzyme PstI, to remove the ORF1b gene. The gene was ligated into the pGEM-3Z vector, which was transformed into E.coli competent cells. The white colonies of E.coli observed on the agar plate were inoculated onto liquid broth medium and incubate overnight. The purified
plasmid DNA was screened by restriction digestion analysis to determine whether the recombinant DNA plasmid contained the insert of the correct size, and whether the orientation is at 5'-3' or 3'-5'sense.

3.5.1 Screening for the presence and orientation of the OFR1b gene

As shown in figure 3.2 A&B (lane 1 to 6), Agarose gel analysis of the restriction digest activities resulted in two bands corresponding the insert and DNA plasmid vector for clones 4 and 6. The insert product obtained, which by comparison to the 1kb DNA marker in lane M, was approximately 1500bp, and 2747bp for pGEM-3Z vector. The band of 423bp confirmed the orientation. Clone 6 was selected and checked again for the insert and orientation as shown in figure 3.3 before proceeding to the next stages of linearization and transcription.

Figure 3.2. A selection of 6 white colonies were analysed for recombinants containing the correct size and orientation of insert using PstI and EcoRI respectively. A) Agarose gel showing the insert, B) Agarose gel showing the orientation of the insert (423bp). Lane M, DNA ladder (1kb DNA), lane 1 to 6, clones containing an insert producing two bands.
3.5.2 Linearization of clone 6 by Nhel, BamHI, and SacI

Clone 6 was linearised separately by Nhel, BamHI, and SacI endonucleases as per fig.3.4. Nhel produced a linearized plasmid DNA containing the T7 promoter as shown by a single band in lane 2A. Hence transcription by polymerase would start from the 5’ end. Similar digestion was carried out using BamHI cleaving the T7 promoter while retaining the SP6 promoter. In this case the DNA was not completely linearised as shown by the presence of two bands as opposed to a single band in lane 3A. Therefore, SacI was used to remove the T7 promoter site and this time as represented by a single band in lane 2B, and which would be transcribed by polymerase from the 3’end, thus producing the required antisense for hybridization in microarray application.
Figure 3.4. Agarose gel showing the linearization of plasmid DNA using Nhel, BamHI, and SacI. A) Lane M: 1Kb DNA marker, Lane 1: uncut clone as a control, Lane 2: digested with Nhel, lane 4: digested with BamHI (not completely cut, as shown by a secondary band of lower intensity). B) Lane M: 1Kb DNA marker, Lane 1: uncut clone, Lane3: digested with SacI.

3.5.3 Determination of transcribed RNA concentration

The in vitro transcription was aimed to obtain copies of RNA in sufficient quantities for labeling. After transcription of the linearized DNA by the MEGAscript kit (Ambion, UK), followed by quantification using the nanodrop. The results for sense and antisense RNA were 2.63, 2.27 µg/µl respectively.

3.5.4 Analysis of pure RNA on gel electrophoresis

The product was run on a gel yielding the bands in figure 3.6. In lane 1, the band of 1500bp was as expected, yet a band of size greater than 1500bp as well as a smearing pattern were also observed. The larger sized and intense band about twice the base pairs of the RNA was possibly due to the formation of some RNA secondary structures that had strangely migrated or due to the migration of plasmid DNA or at least suggesting
impurity or contamination. The smearing pattern might possibly be due to the presence of plasmid components or degradation.

In lane 2, bands of less than 1500bp were observed possibly implying partial or complete degradation. Nevertheless, the RNA concentration obtained was concluded as usable owing to the presence of the band in lane 1 equivalent to 1500bp.

The respective RNA concentrations of 2.63 and 2.27 ug/ul obtained were considered sufficient for labeling (1 ug).

The labeling procedure was carried out as per protocol. The result could only be assessed if and when hybridization took place. Nevertheless, the incorporation of Cy3 and Cy5 needed to be determined to monitor and to ensure the success of the microarray. Before determining the amount of labeled mRNA, it is essential to purify the labeled mRNA, because residual Cy3 or Cy5 direct labeling reagents will interfere with the measurements. This will also minimize hybridization background and improve the sensitivity of detection of low abundance target in microarray.
The incorporation of Cy3 to the positive sense RNA and Cy5 to the negative sense RNA was quantified using UV visible spectrophotometry, nanodrop, ND-1000. Cy3 and Cy5 have absorption maximum at 550 and 650nm respectively. The results were 18 pmol/μl (Cy3), and 3pmol/μl (Cy5).

3.5.5 Analysis of microarray hybridized with labelled RNA

The labeling procedure was carried out as per the manufacturers protocol. The result could only be assessed if and when hybridization took place. Nevertheless, the incorporation of Cy3 and Cy5 needed to be determined to monitor the success of the labeling procedure. Before determining the amount of labeled mRNA, it is essential to purify the labeled mRNA, because residual Cy3 or Cy5 direct labeling reagents will interfere with the measurements. This will also minimize hybridization background and improve the sensitivity of detection of low abundance target in microarray.

The incorporation of Cy3 to the positive sense RNA and Cy5 to the negative sense RNA was quantified using UV visible spectrophotometry, nanodrop, ND-1000. Cy3 and Cy5 have absorption maximum at 550 and 650nm respectively. The results were 2.22 pmol/ng (Cy3), and 2.57 pmol/ng (Cy5).

The transcribed RNAs labeled with Cy3 and Cy5 were pipetted onto the array slide containing the probes and controls. The positive control was the positive sense RNA probes and negative controls were Lucidea Universal ScoreCard controls (Amersham, Biosciences, which have no homology with astrovirus. Upon the addition of the labeled RNA, a reaction arising from hybridization of the negative sense RNA to the probe would be indicated by red colour (Cy5) through hybridization, whilst hybridisation of the positive sense RNA would be indicated by a green color (Cy3). Since the probes were of positive sense this latter was not anticipated although it may arise through non-specific sticking or background. However, once more no colour was detected at all. This could be attributed to the use of lower amounts of labeled samples.

Since this method did not lead to a positive result, a third and final approach was taken to the preparation of labeled reaction products for use in the array. This was the post PCR labeling.
3.6 Analysis of array hybridized with labelled DNA

The primer pair designed (section 3.1.1) was used to amplify known astrovirus RNAs such as serotypes 1, 2, and cloned cDNA by RT-PCR to verify the suitability of those primers. After PCR amplification, a product of 458bp was generated on electrophoresis for HAstV-1, serotype 2, and cloned cDNA, as expected (figure 3.1). No band was observed with negative control (water).

The band from lane 4 was excised and purified as described under section 2.7.4. The pure DNA was denatured and labeled with Cy3 and dispensed on the array for hybridization. It was used as model to investigate if positive hybridization could be obtained. The result of the fluorescent scanning was as shown in figure 3.7.

Figure 3.6. 1.2% Agarose gel of PCR amplification of human astrovirus. Lane 1: negative control. Lane 2: HAstV-1, Lane 3: HAstV-2, Lane 4: cloned cDNA, Lane M: 100 Kb-marker.
The human astrovirus was visually detected by the equal fluorescence emitted on row 4 of blocks 7, 8, 23 and 24 with strongest signal in block 19, 20. It was therefore evident that the system satisfied the specificity requirements for detecting HAstV-1. Obviously, this procedure would seem to offer the best promise in pursuing the screening of viruses in a pan-viral DNA format.
Figure 3.7 Scanning results of hybridization signals on array. Cy3-labelled DNA amplicon was hybridized to oligonucleotide capture probes bound on to glass slide. The variable fluorescent intensities were visualised as green spots on row 4 of block 7, 8, 19, 20, 23, and 24.
4 Discussion

The importance of enteric viruses associated with diseases such as gastroenteritis and hepatitis has been recognised globally. The significance of the Noroviruses in particular has recently become evident and they are now the most commonly reported agents in outbreaks of non-bacterial gastroenteritis (Mead et al. 1999). The survey reported by these authors was based on the compilation of data from multiple surveillance systems. It was estimated that the total illnesses, hospitalizations, and deaths due to enteric virus diseases in the U.S.A approximated to 30.9 million (80%) of the 38.6 million total enteric viral illnesses annually, consequently, making the noroviruses the leading cause of enteric viral illness in the U.S.A. and responsible for 66.6% of the total number of foodborne infections each year (Mead et al., 1999).

However, the diagnosis of viral gastroenteritis worldwide has been drastically underestimated for a number of reasons. Firstly, the lack of sophisticated systems regarding sample collection, analysis and outbreak reporting in developing countries makes the task more difficult in generating accurate or relevant statistical data. Secondly, laboratory methods for the detection of enteric viruses are not always available in the developing world, or conventionally (eg EM) may fail to identify many smaller and indistinct viruses. In recent years, various PCR-based methods have been developed for detection of viral RNA from clinical, food, environmental samples. These assays are rapid, sensitive, and specific. However, they suffer from a problem because of inability to detect different viral pathogens simultaneously (Clewley, 2004). This may be critical in the case of food or water where mixing from a multiplicity of contamination sources may have occurred.

Microarray technology offers greater screening capabilities for the detection and discovery of viruses and an attractive alternative to existing diagnostic methods.

Depending on the availability of appropriate probe sets, microarray can permit the identification of several thousand viruses at the species and subtype level. Although all PCR-based assays described are quite sensitive, specific and rapid they cannot detect many viruses simultaneously. However, the combined application of PCR and microarray
technology for the latter's high-throughput ability can provide a potential and flexible diagnostic assay. The results obtain from the model chosen can be extrapolated to include more probes.

Since gastroenteritis is the condition of interest, the astrovirus was selected as the model for our study and obviously because it is culturable. The strategy followed throughout this project was to first design oligonucleotide probes and primers, and then prepare samples for labeling by harvesting cultures, in vitro transcription and post PCR products.

At present there are two strategies that are available to design microarray, namely cDNA array, and oligonucleotide array. The cDNA array involves linking PCR products to a solid substrate and therefore requires the amplification and purification of thousands PCR probes; a task that would prove very cumbersome. Hence, the less laborious oligonucleotide array was chosen in this study to generate specific probes for microarray detection. Since the primary objective was to detect the viruses only the highly conserved regions were chosen for probe design in this study.

Up to now, the majority of the works on probe design for detection of viruses were carried out based on a set of parameters: a probe length averaging 70 nucleotides (even though 20 nucleotides have been used), and a GC content ranging between 40 to 50% (Li and Stormo, 2001, Wang et al., 2002). The four probes were chosen using a similar set of parameters except that the length of the probe was constrained to 40 bp while the GC content averaged to 42%. The length of 40 bp was selected to maximise specificity since longer probes tended to average out similarity between all astroviruses because this region was so well conserved. In addition 40 bp was selected instead of shorter ones, because the plan was to evaluate the detection of viruses in environmental samples, which no one has done yet. Once positive results are obtained the idea was to aim for shorter probes to detect viruses in food and water samples. At this length hybridisation took place with the target. However, such positive results could have been improved for sensitivity and specificity had we known about the importance of Tm, which no one appeared to have mentioned. The Tm should normally range between $80^\circ C$ to a maximum of $85^\circ C$ for shorter probes (Baxi et al, 2005). Assuming the rule of thumb for Tm to primers can be applied to probe design then the Tm of the probe with a GC content averaging 42% will always be on the higher side, giving an approximate value of $110^\circ C$ as opposed to a maximum of $85^\circ C$. Therefore, it seems possible that the restriction on G/C content could
be reduced to a lower percentage. For most 70 nucleotide probes that gave positive results, it appears that the Tm was probably not necessarily an essential requirement.

Thus, despite the positive results, better probes with perhaps shorter lengths can be designed when the Tm parameter is taken into consideration.

At the time of designing the probes information regarding parameters such as those governing secondary structure formation and self-complementarity and cross hybridisation were not readily available. Since these parameters were not included in the design it is quite reasonable that when they are taken into consideration better probes can be expected.

The homology is also as important as the length of the probe. This is because of the possibility of cross hybridization if the non-target sequence has more than 75-80% similarity with the probe.

Overall, when considering all parameters in context and evidently with the advent of more sophisticated programs or algorithms, cheaper probes with better sensitivity and specificity can be designed, thus enhancing the efficacy of the probes.

Similarly, primers were designed using the same criteria as for the probes except in this case the Tm had a range varying from 50-74°C. The result as depicted in figure 3.6 suggests that the primers have worked. This was confirmed by the amplification of the target sequence shown as a single band of 458bp, as expected. The intention was to use these primers for post-PCR labelling in house. Since most of the available labelling kits come with their own primers, the designed primers were aliquoted and frozen in the event that the cDNA and mRNA labelling techniques might not work and that was the case.

There are several ways of labelling RNA or DNA from samples to be used for analysis by microarray. They are the direct (cDNA, mRNA, PCR amplicons, genomic DNA, plasmid DNA), and indirect approaches. However, the focus was mostly on the direct approach that would appear more logical.
Hence, the choice was to directly label the first strand cDNA. However, the results were not conclusive because of the low levels of RNA obtained. Even though the minimum amount of total RNA required was 2.5μg it was quite difficult to label the RNA, leading to no results. It seems that most researchers use a concentration of 10 to 30 μg of total RNA (Lee et al., 2003; Bystricka et al., 2005). Despite scaling up and protecting the RNA no positive results were obtained possibly due to RNA degradation. Another reason for the failure of this experiment could have been due to the use of oligo-dT as the primer for the amplification of the target (orf1b) that is about 2360bp from the 3’ end. Theoretically, it is possible yet it did not work. Other workers have employed the oligo-dT approach but they found that the poly (dA/dT) was largely responsible for the low efficiency of identification of the genes expressed at lower levels (Wang et al., 2000; Nam et al., 2002). This seems to explain our observation. The other approach would have been to use the random nonamer for pure RNA, as per protocol (CyScribe, Amersham), but since the yield of RNA was too low it was not worthwhile because of the risk of further loss. Other researchers have used random primers successfully to amplify their targets; the advantage of this method is that no prior knowledge of the infectious agent is required yet larger amounts of total RNA were still required (Wang et al., 2002). Therefore, the second approach was to directly label the purer RNA after it had been produced by transcription in vitro.

Cloning would not be practical in screening procedures but was useful for array evaluation. Of the six positive clones selected, only two were used for further experiments because the remaining four showed incomplete digestion (fig 3.2A lanes: 1, 2, 3, and 5). Following linearization and transcription of the two clones, the concentrations of RNA obtained were within the minimum requirement of 1 μg.

Despite the benefit of a one-step non-enzymic method without reverse transcription or amplification, the use directly labelled mRNA failed to yield positive results as indicated by the lack of signals. The problems were possibly due to partial degradation, contamination or impurities, as shown by two bands including a smear as opposed to a single have led to further loss of RNA. Nevertheless, no results were derived. Since RNA is relatively unstable compared to DNA, the last approach was post employed was PCR labelling “genomic DNA labelling”.
It appeared that post PCR labelling provided a good means for the detection of HAstV-1 using Cy3 dye by microarray. The reasons could have been due to the fact that the amplified target was produced in sufficient quantity and more stable.

Even though the qualitative results were good, subjective factors could be associated in the interpretations of the outcomes. Therefore, for more precise interpretations improved versions of analytical software may be used. In addition, some experiments have been carried out incorporating software algorithms that check for secondary structures and normalize melting temperatures. Nevertheless, a number of probe failures were encountered whereby a target has not hybridized to a probe as expected irrespective of probe length. It seems that secondary structure has a major role in reducing hybridisation or fluorescence intensity (Lane et al., 2004).

Consequently, based on what is known relative to the genomic labeling experiment conducted, a viable suggestion would be to design shorter probes of lengths ranging in sizes from 10 to 15 bp and direct labeling of target in one step. Direct labeling has the advantage of being time effective but may compromise detector sensitivity. Nick translation is one means of disrupting secondary structure and does not compromise detector sensitivity.

Besides, there are several approaches for target labeling that can be selected. Each one has its own merits and obviously dependent on the cost. Nonetheless, the ideal recommendation would probably be to use PCR amplicons because the procedure involves a single step and specific or random primer sets can be utilized.

Unfortunately, further experiments for the complete evaluation of the array could not be conducted owing to time limitation. Nevertheless, microarray still remains a powerful tool for rapid and potentially cost effective diagnosis of multiple pathogens in a single test. Another advantage of microarray is that it has 30-fold sensitivity more than agarose gel electrophoresis for detecting a short PCR product.
In conclusion, microarray in conjunction with post PCR labelling would probably offer a better option for the detection and genotyping of viruses to achieve the objective set up at the beginning.

However, when designing a microarray assay, all relevant factors, specially sample preparation need to be considered with respect to improving assay sensitivity. This is because food or environmental samples generally produce low template yields.

Future work

- To continue evaluating the post PCR labelling method
- To establish and validate microarray-based methods for the simultaneous detection of multiplicity of viruses
- To use this method in clinical and foodborne contexts.
5 Appendix

General Solutions

TBE Running Buffer (10X): 1 litre
107.8 g Tris base
55 g Boric acid
9.3 g EDTA (diNa.H₂O)
Made up to 1L with sterile distilled water. Used in 1x concentration as running buffer and as a means to dissolve the Agarose.

TBE Sample Buffer
4 mls 10X TBE
4 mls glycerol
12 mls distilled H₂O
Add small amount Bromophenol blue dye and Xylene Cyanol FF dye.

Sodium Acetate 3M
Sodium Acetate MW = 82.05 g/mol
24.61 g of sodium Acetate was dissolved in 50 µl of distilled water and stirred until dissolve.
Adjust pH to 5.2 with glacial acetic acid
Made up to 100ml
Sterilise by autoclaving
Appendix

**Luria-Bertani (LB) broth**: 1 litre

- 10g Tryptone
- 5g Yeast extract
- 10g NaCl
- Adjust pH to 7.5 with 5M NaOH
- Sterilise by autoclaving

**Luria-Bertani (LB) agar**: 1 litre

Same as LB broth (above) plus 1.5% agar (w/v) (BACTO™ Agar).
- Sterilise by autoclave.

**Addition of antibiotics for selective growth**

- pGEM-3Z+ ORF1b transformation
- AMPICILLIN (25mg/ml): 1/500
- 20% X-Gal: 75-100μl in 100ml
- IPTG (100Mm): 1/400

**IPTG stock solution (0.1M)**

Nuclease free water was added to 1.2g of IPTG (Promega Corporation, Madison) to a final volume of 50 ml. The IPTG solution was filter sterilised and stored at 4°C.

**X-Gal stock solution**

100 mg of 5-bromo-4-chloro-3-indolyl-p-d-galactoside (Promega Corporation, Madison) was dissolved in 2 ml N, N’-dimethyl-formamide. The solution was covered in aluminium foil and stored at −20°C.
2.5 M NaOH for 10ml
1 g NaOH
Nuclease free water to 10 ml
Sterile by filtration with a 0.45-micron filter
The solution was stored at room temperature for up to 3 months

2M HEPES free acid for 10ml
4.77 g HEPES free acid
Made up to 10ml nuclease free water

Phosphate buffered saline (100 tablets - Dulbecco)
1 tablet was dissolved in 100 ml of distilled water
Sterilise by autoclave at 115°C for 10 min
6 References


References


References


*J. Pediatr.* **120**: 516-521.


*Arch. Virol.* **113**: 73-81.


*FEMS Microbiol. Lett.* **114**: 1-7


