Detection and Quantification of Variable Viral RNA by Real-Time PCR Assays

SHAMAN MURADRA SOSOLI
Dissertation presented at Uppsala University to be publicly examined in Hörsalen, Baktlab ing D1, Dag Hammarskjölds väg 17, Uppsala, Friday. September 5, 2008 at 13:00 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in Swedish.

Abstract

As the area of nucleic acid-based technologies develops, so will our understanding of how structural variations in DNA and RNA pathogens are associated with disease. The overall goal of this thesis is the development of broadly targeted measurement techniques for variable viral RNA by Real-Time PCR (here referred to as quantitative reverse transcriptase PCR, QRT-PCR).

In papers I & II, broadly targeted and specific QRT-PCRs were used to study expression of endogenous and exogenous betaretrovirus sequences in human tissues. Results from human tissues demonstrated endogenous betaretrovirus expression in a tissue-specific manner, highest in reproductive tissues. Despite the high sensitivity, no exogenous betaretrovirus was found in human breast cancer samples. The limits of primer and probe degeneracy for detection of a diverse set of retroviral sequences was evaluated. These methods are useful for further investigations on the pathophysiological contribution(s) of endogenous betaretrovirus and to investigate whether an exogenous betaretrovirus is involved in human breast cancer.

In papers III & IV, we developed and applied broadly targeted one-step QRT-PCRs for influenza viruses and coronaviruses. In addition to the generic primers, two novel probe design strategies were used in order to be able to broadly amplify these diverse sets of viruses: A triplex system for simultaneous detection and quantification of influenza A, B and C (3QRT-PCR and further developed 3QRT-PCR-MegB; where MegB stands for MegaBeacon) based on TaqMan® and MegB probes, and a pan-CoV QRT-PCR, based on three TaqMan® probes i.e., degeneracy was distributed on three probes. Probe fault tolerance was thus increased in two ways, either with short probes with/without locked nucleic acid (LNA) nucleotides concentrated to conserved stretches, or with long probes (MegB), compensating mismatching positions with many matching ones. Clinical samples, negative by antigen detection with immunofluorescence (IFA), were influenza A positive with 3QPCR-MegB. Avian pooled samples, negative with an earlier pan-CoV QPCR, came out positive with the triple-probe system. Assay evaluation with clinical samples and reference strains revealed good clinical diagnostic potential.

Thus, the thesis describes several strategies to counteract sequence variation of RNA viruses and describes a set of broadly targeted QRT-PCRs useful for scientific screening or diagnostics of betaretroviruses and respiratory viruses.

Keywords: Virology, Diagnosis of Viral RNA, QPCR, RNA virus, broadly targeted PCR, probe design strategy, Infectious diseases

Shaman Muradrasoli, Department of Medical Sciences, Clinical Virology, Akademiska sjukhuset, Uppsala University, SE-75185 Uppsala, Sweden

© Shaman Muradrasoli 2008

ISSN 1651-6206
ISBN 978-91-554-7251-1
urn:nbn:se:uu:diva-9193 (http://urn.kb.se/resolve?urn=nbn:se:uu:diva-9193)
In memory of my father
Who always told me to be curious and that questioning leads to more learning.
List of papers

This thesis is based on the contribution of the author to papers I to IV indicated below. They will be referred to by their Roman numerals:


III. Muradrasoli S\textsuperscript{1}, Mohamed N\textsuperscript{1}, Belák S, Czifra G, Herrmann B, Berenksi G, Blomberg J. Broadly targeted triplex real-time PCR detection of influenza A, B and C Viruses based on the nucleoprotein gene and a novel probe strategy. (Submitted, revised)

IV. Muradrasoli S, Mohamed N, Hornyák Á, Fohlman J, Belák S, Blomberg J. Broadly targeted triple-probe QPCR for detection of coronaviruses. Coronavirus is common among mallard ducks (\textit{Anas platyrhynchos}) (Submitted, revised)

\textsuperscript{1} These two authors contributed equally, and share the first authorship.
\textsuperscript{2} Reprints were made with permissions from the publishers. \textsuperscript{\textcopyright} 2006 Elsevier B.V.
Supervisors and Opponents

Main supervisor

Professor Jonas Blomberg
Department of Medical Sciences,
Section of Clinical Virology
Academic Hospital
Uppsala

Co-supervisor

Professor Göran Friman
Department of Medical Sciences
Section of Infectious Diseases
Academic Hospital
Uppsala

Opponent

Docent Maria Brytting
Swedish Institute for Infectious Diseases control
Section of Virology
Stockholm
Broadly targeted QPCR ................................................................. 33
  Design of degenerated primers .................................................... 33
  Design of degenerated probes .................................................... 34
  Introduction of LNA™ residues to enhance \( T_m \) ....................... 35
Novel probes: Triple probe and MegaBeacon ............................. 36
  Triple-probe concept .................................................................. 36
  MegaBeacon probe .................................................................... 37
  Comparison of MegB and conventional MB ............................ 38
Further PCR optimization .............................................................. 38
  Search for representative sequences ......................................... 40
  Optimisation of annealing temperature ..................................... 40
  Multiplex QRT-PCR ................................................................. 40
  Touch-down QRT-PCR ............................................................. 41
  Amplification efficiency ............................................................ 41
  Sensitivity and specificity .......................................................... 42
  Stochastic zone in quantitative PCR ......................................... 42
  Contamination precaution ........................................................ 43
Results and Discussion ................................................................. 44
  Endogenous MMTV-like sequences in human tissues, Paper I .... 44
  Exogenous MMTV-like sequences in breast cancer, Paper II ...... 47
  Simultaneous detection of influenza types A, B and C, Paper III .... 49
  Detection of coronaviruses in humans and animals, Paper IV ...... 51
Concluding remarks ...................................................................... 53
Acknowledgments .......................................................................... 55
References ...................................................................................... 58
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>capsid protein</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>env</td>
<td>envelope gene</td>
</tr>
<tr>
<td>FRET</td>
<td>fluorescence resonance energy transfer</td>
</tr>
<tr>
<td>HML</td>
<td>human mouse mammary tumor virus like</td>
</tr>
<tr>
<td>HERV</td>
<td>human endogenous retrovirus</td>
</tr>
<tr>
<td>HERV-K</td>
<td>human endogenous retrovirus K</td>
</tr>
<tr>
<td>ICTV</td>
<td>international committee on taxonomy for virus</td>
</tr>
<tr>
<td>IN</td>
<td>integrase</td>
</tr>
<tr>
<td>LTR</td>
<td>long terminal repeat</td>
</tr>
<tr>
<td>MA</td>
<td>matrix protein</td>
</tr>
<tr>
<td>MB</td>
<td>molecular beacon</td>
</tr>
<tr>
<td>MegB</td>
<td>mega beacon</td>
</tr>
<tr>
<td>MMTV</td>
<td>mouse mammary tumour virus</td>
</tr>
<tr>
<td>NC</td>
<td>nucleocasid</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>pol</td>
<td>polymerase gene</td>
</tr>
<tr>
<td>pro</td>
<td>protease gene</td>
</tr>
<tr>
<td>QPCR</td>
<td>quantitative polymerase chain reaction</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RT</td>
<td>reverse transcriptase</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>reverse transcriptase polymerase chain reaction</td>
</tr>
<tr>
<td>SU</td>
<td>surface protein</td>
</tr>
<tr>
<td>TM</td>
<td>transmembrane protein</td>
</tr>
</tbody>
</table>
Introduction

Background
Given that most of the evolution of life on this planet has occurred in the sea, viruses of nearly all the major classes of organisms - animals, plants, fungi and bacteria/archaea probably evolved with their hosts there. This means that viruses also probably emerged from the waters with their hosts, during the successive waves of colonisation of the earthly environment (Forterre 2006). Viruses, microscopic particles that are obligate intracellular parasites, can infect the cells of biological organisms. Viruses replicate themselves by infecting a host cell; once it has entered the host cell and it hijacks the cellular machinery to facilitate the viral replication and promulgation. Viruses consist of genetic material enclosed within a protective protein coat (Sauer 1970). All organisms, with the exception of the RNA viruses, encompass their genetic material in form of DNA, using RNA only as a temporary messenger for information. DNA is a quite stable molecule, not particularly reactive with other molecules. DNA dependent DNA polymerases make very few errors during replication, the processes of reproduction of DNA (between one in 1 million and 1 in 10 million) (Alberts 2008). Most of these mistakes are normally quickly corrected even when they do occur. This makes DNA an ideal format for the storage of information, for mutations (errors) only rarely occur, and most are not significant. RNA, by contrast, is a quite reactive molecule, capable of reacting even with itself under the correct conditions. High mutation frequencies of RNA dependent RNA polymerases of RNA viruses have been observed since earliest genetic studies (Granoff 1961; Fields and Joklik 1969; Pringle 1970). RNA polymerase makes frequent mistakes during copying - averaging one mistake per 10,000 nucleotides each time it is copied (Alberts 2008). However, these unique aspects give RNA advantages for the storage of viral information. The immune system develops immunity by learning to recognize an infecting virus and creating antibodies against it. Once antibodies are produced and immunity is developed, the virus is quickly destroyed, so the virus can no longer use that host for reproduction. Some viruses, like many RNA viruses, must change their nature so that the immune system will no longer recognize them, in order to reinfect a host. In other words, it must mutate.
Genes that are not or are very little affected by mutations are regarded as conserved. Conserved genes are likely to play an essential role in the virus lifecycle and are therefore primary targets for diagnostic methods. There are several potential advantages of targeting the most conserved features of the virus genomes: effective inhibition of viral growth, decreased rates of drug resistance, fewer drugs needed and broad spectrum drugs.

Virus structure and classification

Viruses can be classified in several ways, such as particle morphology, genome type, replication strategy, host organism, mode of transmission, or by the type of disease they can cause. The two most commonly used classification systems are the Baltimore classification and the International Committee on Taxonomy of Viruses (ICTV) system. The Baltimore system is based on replication strategy, i.e. the nature of the virus genome and its mode of expression and classifies the viruses into seven groups: Double-stranded DNA (ex. Adenoviruses, Herpes viruses, Poxviruses), Single-stranded (+) sense DNA (ex. Parvoviruses), Double-stranded RNA (ex. Reoviruses, Birnaviruses), Single-stranded (+) sense RNA (ex. Picornaviruses, Togaviruses), Single-stranded (-) sense RNA (ex. Orthomyxoviruses, Rhabdoviruses), Single-stranded (+) sense RNA with a DNA intermediate (ex. Retroviruses), and Double-stranded DNA with a RNA intermediate (ex. Hepadnaviruses). The ICTV system is based upon the Linnaean System of Taxonomy and employs the hierarchical taxonomic levels of order (-virales), family (-viridae), subfamily (-virinae), genus (-virus) and species (species names can contain more than one word and have no specific ending) (Table 1). Based on this system there are 3 orders, 56 families, 9 subfamilies, and 233 genera. ICTV recognizes about 1,550 virus species but virologists are tracking about 30,000 virus strains and isolates.

Table 1. The seven taxonomic levels of the ICTV system

<table>
<thead>
<tr>
<th>Taxonomic level</th>
<th>Suffix</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order</td>
<td>-virales</td>
<td>Mononegavirales</td>
</tr>
<tr>
<td>Family</td>
<td>-viridae</td>
<td>Paramyxoviridae</td>
</tr>
<tr>
<td>Subfamily</td>
<td>-virinae</td>
<td>Paramyxovirinae</td>
</tr>
<tr>
<td>Genus</td>
<td>-virus</td>
<td>Rubulavirus</td>
</tr>
<tr>
<td>Species</td>
<td>(a single kind of virus)</td>
<td>Mumps virus</td>
</tr>
</tbody>
</table>
RNA viruses

RNA is a polymer made from repeated units called ribonucleotides covalently linked together. The chemical differences between DNA and RNA are that the ribose in RNA is hydroxylated and the pyrimidine base thymine is replaced by uracil. RNA viruses and retroviruses use RNA as genetic material, which means the RNA serves as genome template and not as an intermediate template for translation of genes into proteins. Many RNA viruses show a great sequence variability that is partly caused by point mutations introduced into the viral genome during replication due to the low fidelity of RNA dependent RNA/DNA polymerases. The DNA viruses have a low mutation rate as a result of polymerase proofreading activity, while no such error correction is present in RNA dependent RNA/DNA polymerases (Smith and Inglis 1987; Steinhauer, de la Torre et al. 1989; Pathak and Temin 1990; Pathak and Temin 1990).

RNA viruses comprise 350 different human pathogens. They are responsible for many emerging diseases. Effective medical treatment of RNA virus infection is a challenge for researchers, since many RNA viruses need only a few weeks to escape the immune response or to evolve drug resistance. Properties that contribute to variation in the virus population are a short generation time and a small genome, i.e. every error or mutation can contribute to a phenotypical change. Alteration of nucleic acid sequence can also occur through recombination, which is exchange of genetic information between existing variants. New variants are generated through recombination when two different viruses coincidentally infect the same cell and their genes are mixed, and a third variant is produced with features of the two. The obvious consequence of genetic diversity resulting from point mutation, is that RNA viruses possess a great potential for rapid evolution and adaptation (Lai 1992).

Replication strategies of RNA viruses

Since the host cell does not contain RNA dependent RNA polymerase, the replication of viral RNA genomes is unique in that the majority of RNA viruses encode their own RNA polymerase to overcome this obstacle. Some viruses pack their polymerase into the virus particle, others synthesize after infection. The genome of RNA viruses exists in a number of configurations, including unimolecular or segmented, positive-stranded, negative-stranded, double-stranded or circular. In the positive-stranded RNA viruses, the virion (genomic) RNA is the same sense as polycistronic mRNA. The viral mRNA is translated immediately upon infection of the host cell into a polyprotein product. The mature proteins are formed from cleavage of the polyprotein.
Coronaviridae, flaviviridae and picornaviridae are examples of this class. The negative-stranded RNA viruses can have either segmented or non-segmented genomes. In both cases, the first step in replication is to transcribe a sense RNA genome to produce a monocistronic mRNA, which subsequently serves as the template for genome replication. Orthomyxoviridae and paramyxoviridae, are examples of this class. Replication of double-stranded RNA viruses is monocistronic, and they have segmented genomes that are transcribed to mRNAs. Reoviridae and Birnaviridae are examples of this class (Fields, Chanock et al. 1996).

Conserved secondary structures in viral RNA genomes

Almost all RNA molecules form secondary structures. The genomes of RNA viruses not only code for proteins, but often contain functionally active RNA structures. Functional secondary structures evolve much slower than the underlying sequences, and have most likely an important function in the viral life cycle (Fontana, Konings et al. 1993). In general, functionally active secondary RNA structures are susceptible to point mutations, and do not differ from the ones formed by random sequences: computer simulations reveals that large changes in the secondary structures come up from a small number of point mutations (Fontana, Konings et al. 1993; Schuster, Fontana et al. 1994). Viruses use different approaches to ensure their genome amplification. This can be achieved in an efficient and in some cases, a cell-type specific manner, the majority of RNA viruses encode their own polymerases which discriminatory amplify the viral genomes; other viruses have evolved a variety of ways to compete directly with the host cell for factors that are needed for viral gene replication and packaging. During early stages of infection viruses have to compete with the host for both the ribosome and for the limited pool of eukaryotic initiation factors (eIFs) that are needed to facilitate the recruitment of ribosomes to both viral and cellular mRNAs. Some viruses have evolved to recruit ribosomes by cis-acting elements, known as the internal ribosome entry site (IRES). The IRES elements were first described within the RNAs from various picornaviruses including poliovirus, human rhinoviruses, encephalomyocarditis virus and foot-and-mouth disease virus (Jang, Krausslich et al. 1988; Pelletier and Sonenberg 1988). The IRES is a highly specialized RNA element that is often located at the 5’end of the viral genomes. It allows assembly of the translational machinery at a position close to or directly at the initiation codon without a 5’ cap structure. IRES is conserved and is therefore a useful target for broadly targeted PCRs.
Pathogenicity and virulence of RNA viruses

When the terms pathogenicity and virulence are applied to viruses, it is important not to use them without reference to the host. These terms always refer to the interaction between the virus and its infected host. Pathogenicity of a virus is its ability to infect and induce disease in a host (Watson 1949; Smith 1977). Virulence, on the other hand, is the severity of the disease after infection has occurred. Virulence is measured by case fatality rate or proportion of clinical cases that develop severe disease (Youmans 1975; Wood 1980). Pathogenic steps of viral disease include: implantation of virus at the portal of entry, local replication, spread to target organs (disease sites), and spread to sites for viral shedding into atmosphere. The pathogenic mechanisms can be affected by factors such as: accessibility of virus to tissue, cell susceptibility to virus multiplication, and virus susceptibility to host defenses. A number of RNA viruses are highly infectious, acutely pathogenic, and are a significant source of morbidity and mortality in humans every year. Newly emerging RNA viruses may endanger public health.

Transmission of enveloped RNA viruses

Enveloped viruses interact with their host cell through specific binding interactions between glycoproteins on their envelope and receptors on the cell surface. They have to pass through the plasma membrane of the host cell twice during the viral replication cycle: initially during virus entry and again during particle release (Freed 2004). Entry is typically achieved by means of a membrane fusion reaction, occurring either directly at the cell surface following particle binding, or in low-pH endosomes after endocytosis of bound virions (Kielian and Jungerwirth 1990; Barocchi, Masignani et al. 2005; Chou 2007). Enveloped viruses have evolved a diversity of strategies for release from the cell (Freed 2004). Common pathways of virus entry include skin, respiratory tract, gastrointestinal tract, genitourinary tract and conjunctiva. Most RNA viruses transmit via the fecal-oral route and/or the respiratory tract, where they initiate infection by replicating in epithelial cells (Compans 1995; Bomsel and Alfsen 2003). Respiratory infection may be initiated either by virus contained in aerosols that are inhaled by the recipient host or by virus that is contained in nasopharyngeal fluids and is transmitted by hand-to-hand contact. Viral infections can be localized to the site of inoculation. However, virus propagation in the blood is the most important route since viruses can potentially be carried to any site in the body. This is known as viraemia (Howley 2001). The approach of transmission varies for each virus. For instance, retroviruses integrate their genome into host DNA and when integrated into germline cells it becomes endogenized, e.g. endogenous retroviral sequences (ERVs). Understanding the routes of entry and modes of transmission is important for selection of the right sample material.
RNA viruses as causes of zoonoses

A zoonotic pathogen causes an infection in animals which can be transmitted to humans. The most serious zoonoses are often viral in origin, especially from RNA viruses with their ability to adapt to changing environmental conditions. They can cause severe disease when transmitted to humans from an animal reservoir (Bray 2008). As animals have been domesticated and the human-animal interaction is of increasing closeness, the risk of zoonotic diseases increases. The emergence of zoonotic viruses maintained by wildlife reservoir is poorly understood. Learning more about zoonotic diseases has become increasingly important during recent decades. The majority of the emerging infectious diseases within the last 25 years have had zoonotic origins. The risk of zoonotic infection shows no sign of diminishing and could increase in the future. Approximately 75% of recently emerging infectious diseases affecting humans are of animal origin, both domestic and wildlife (Bengis, Leighton et al. 2004; Brown 2004; Kuiken, Leighton et al. 2005). HIV type 1 (HIV-1), the major cause of AIDS, is related to SIV found in chimpanzees (Gao, Bailes et al. 1999). Our protection against viral zoonoses depends on the ability to identify unusual outbreaks of disease as early as possible, by development of rapid and sensitive detection methods.

Human RNA viruses in the present work

The retrovirus family (Retroviridae)

The retrovirus family (Retroviridae) is classified into seven genera; alpha-, beta-, gamma-, delta- and epsilon-, lenti- and spumaretrovirus. A retroviral genome consists of a single positive stranded RNA molecule, with a size of 7-10 kb. The general structure of all retroviral genomes is that the proviral DNA strand consists of at least four genes. The integrated viral genome (the provirus) is bracketed by long terminal repeats (LTRs). The two LTRs are identical in sequence and organization. These repeats contain regulatory elements such as enhancers and promotores, which regulate the expression of the viral genes. The promoter elements in the U3 regions are responsible for the initiation of transcription. The genes that build up the structure of the new virus particles present in all retroviruses are; gag (group antigen proteins), pro (protease), pol (polymerase) and env (envelope); The gag gene encodes structural proteins to form the virion, which includes the proteins making up the matrix (MA), the capsid (CA) and the nucleoprotein (NC); the pro open reading frame (ORF) encodes an enzyme that cleaves the gag polyprotein precursor; pol encodes reverse transcriptase (RT) and integrase (IN) which is involved in provirus integration (see below); env encodes the surface glycoproteins(SU) and transmembrane (TM) protein (Hunter and Swanstrom 1990).
Human endogenous retroviruses (HERVs)

The human genome consists of many retroviral elements that are homologues of known animal exogenous retroviruses. Human endogenous retroviruses (HERVs) comprise up to 8% of the human genome (Lander, Linton et al. 2001). Retroviruses are enveloped viruses that rely on the enzyme reverse transcriptase to perform the transcription of its RNA into DNA, which is integrated into the host cell genome with the help of the enzyme integrase. After this integration the virus is called a provirus. The provirus uses the host cell machinery for replication and new virus particles are produced to infect other cells (horizontal spread). A retrovirus becomes endogenous when a retrovirus particle infects germ cells. Once an endogenous retrovirus has been integrated and established in the germ line it is inherited from generation to generation (vertical spread). HERVs have been inherited by successive generations and it is possible that some have conferred biological benefits. However, several HERVs have been implicated in certain cancers and autoimmune diseases (Stoye and Coffin 1985; Blomberg, Ushameckis et al. 2004).

HERVs are often divided into three main classes - I, II and III, where class I are gammaretrovirus-like, class II are betaretrovirus-like and class III are distantly spumaretrovirus-like (Medstrand and Blomberg 1993). The nomenclature for HERV is complex and there is no well-defined system. The classification of betaretroviruses defined by the ICTV (Fauquet and Martelli 1995) does not yet encompass endogenous retroviral sequences. HERVs were earlier classified according to the tRNA primer-binding site (PBS) they have in the 5’ leader region. A general treatise of retroviral taxonomy was given by (Jern, Sperber et al. 2005). The human betaretrovirus-like sequences were originally identified by their sequence similarity to MMTV. The human betaretroviral PBS is often complementary to lysine tRNA and the group is therefore often called HERV-K. To distinguish groups of elements similar to MMTV, the HERV-K group is divided into ten subgroups, HML1-10; where HML stands for “human MMTV like” (Medstrand, Lindeskog et al. 1992; Andersson, Lindeskog et al. 1999; Johnson and Coffin 1999; Gifford and Tristem 2003; Lavie, Medstrand et al. 2004). HERV-K(hml2) is considered to be the HERV with greatest potential for biological activity because some members of this group have intact open reading frames for the gag, pol or env genes (Tonjes, Lower et al. 1996). In addition to the structural proteins, some of the HML groups may produce regulatory proteins, such as rec and Np9; both are expressed from spliced mRNAs from the env reading frame by alternative splicing. The function of Np9 is unknown, although it is expressed in a variety of tumors and transformed cell lines (Lower, Lower et al. 1996; Magin, Lower et al. 1999; Armbruester, Sauter et al. 2002; Armbreuster, Sauter et al. 2004). Even though the bio-
logical function of HERVs has never been fully elucidated, HERVs have been associated with different diseases. Attempts to associate HERVs with disease show that they might have a possible involvement in the development of autoimmunity and multiple sclerosis (Rasmussen, Geny et al. 1995; Christensen, Sorensen et al. 2001; Christensen, Pedersen et al. 2002; Clausen 2003).

Mouse mammary tumour virus (MMTV) and related retroviruses

Mouse mammary tumour virus (MMTV) is the prototype betaretrovirus and causes breast cancer in mice, by insertional mutagenesis. It was discovered in 1930 by Dr. John J. Bittner (Bittner 1936). There are two routes of transmission of the virus in mice, i.e. transmission of an exogenous or endogenous form of the virus. In the first route, the virus is transmitted during lactation since it is present in the milk. The endogenous form is transmitted via vertical spread from parent to offspring. It was recently demonstrated that human endogenous retroviral sequences distantly related to MMTV, HERV-K(hml2) could be "repaired". Two research groups reconstructed infectious HERV-K(hml2). Gag and protease proteins of the HERV-K clone were able to mediate assembly and processing into retrovirus-like particles. The capability of infectious transfer of the construct and stable integration was verified by insertion of reporter genes. (Dewanneix, Blaise et al. 2005; Lee and Bieniasz 2007). Beside this, HERV-K LTRs contain signals for transcription initiation as well as termination and cleavage polyadenylation (the process of eukaryotic mRNA 3' end formation, is essential for gene expression and cell viability) found in retroviral LTRs, i.e. a TATA-box and typical polyadenylation signal (AATAAA) (Ono 1986).

MMTV can subvert parts of the immune system of the infected mouse. The superantigen (SAg) is encoded from an open reading frame within the 3' long terminal repeat. This interesting feature, which is unique to MMTV, facilitates viral replication and enhances infectivity, via interaction between viral SAg, T-cell receptor and MHC class II I-E. MMTV utilizes cells of the immune system in its infection pathway from infected milk in the gut, to target cells in mammary glands. Mice that retain non-productive endogenous proviruses are generally considered to be protected from infection with exogenous virus (Golovkina, Chervonsky et al. 1992; Held, Shakhov et al. 1993; Held, Waanders et al. 1993; Beutner, Kraus et al. 1994). Coexpression of exogenous and endogenous MMTV RNA could also be advantageous for the virus, by increasing its variability, broadening the host range and allowing the expansion of highly tumorigenic variants (Golovkina, Jaffe et al. 1994).
Influenza viruses, the *Orthomyxoviridae*

Influenza viruses is a negative sense single stranded RNA, which belongs to the Orthomyxoviridae family. They are characterized by a high mutation rate resulting in antigenic drift and a high frequency of genetic rearrangement resulting in antigenic shift (Hilleman 2002). These properties lead to remarkable changes in the antigenicity of the viral surface glycoproteins. There are three types of influenza viruses, type A, B and C, classified by antigenic differences in viral nucleoprotein and matrix proteins. The RNA of influenza viruses is segmented into 8 distinct genes (7 for types C), coding for ten different proteins. Influenza A is divided further into subtypes based on antigenic differences of the surface proteins hemagglutinin (HA) and neuraminidase (NA) (Webster, Bean et al. 1992). Currently, there are 16 distinct HA (H1 to H16) and 9 NA (N1 to N9) subtypes (Fouchier, Munster et al. 2005). Influenza A displays the highest genetic variability of the three influenza types. It largely results from antigenic drift, which allows the virus to escape antibody neutralization. In antigenic shift two influenza A viruses combine to form a new subtype by exchange of genomic segments. The new subtype has a mixture of segments from two viruses.

The influenza A virus causes influenza in birds and some mammals, and these influenza viruses occur naturally among birds. Wild anseriform birds (e.g. ducks) worldwide carry the viruses in their intestines, but usually do not get sick from them. Such lowly pathogenic (LPAI) viruses can become highly pathogenic (HPAI) avian influenza viruses either by insertion, mutation or recombination (Ludwig, Stitz et al. 1995; Perdue, Garcia et al. 1997; Suarez, Senne et al. 2004; Pasick, Handel et al. 2005). If highly pathogenic strains are transmitted from wild aquatic birds to domestic poultry, including chickens, ducks, and turkeys, they can make them very sick and kill them. During an outbreak of avian influenza among poultry, there is also a risk to people who have contact with infected birds, or with surfaces that have been contaminated with secretions or excretions from infected birds (Bridges, Kuehnert et al. 2003). This could potentially cause a major outbreak in humans, in the worst case a new human influenza pandemic. However, aside from relatively few incidents this has not occurred. “Human influenza virus” usually refers to those subtypes that spread widely among humans. There are only three known A subtypes of influenza viruses (H1N1, H1N2, and H3N2) currently circulating among humans. Three pandemics have occurred during the 20th century, caused by three different influenza A subtypes, Spanish flu, Asian flu and Hong Kong flu (Potter 2001; Taubenberger, Reid et al. 2001) (Table 2).
Table 2. Influenza pandemics occurred in the 1900s

<table>
<thead>
<tr>
<th>Name</th>
<th>Time period</th>
<th>Subtype</th>
<th>Deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish flu</td>
<td>1918</td>
<td>H1N1</td>
<td>40-50 million</td>
</tr>
<tr>
<td>Asian flu</td>
<td>1957</td>
<td>H2N2</td>
<td>2 million</td>
</tr>
<tr>
<td>Hong Kong flu</td>
<td>1968</td>
<td>H3N2</td>
<td>1 million</td>
</tr>
</tbody>
</table>

Influenza B and C show lower genetic variability than influenza A. Up to now, the mechanisms responsible for changes in these viruses are not well characterized (although it is likely to exist). Although, genetic variability of influenza A and B is higher there has been reports showing that influenza C undergoes genetic changes like influenza A or B (Goto, Tanaka et al. 1984). Influenza A and B can co-circulate, and be a major source of serious illness, hospitalisation and death in elderly persons (Hampson and Mackenzie 2006). Although, influenza C was isolated more than 60 years ago, there are few reports describing its clinical features. Influenza C is associated with common cold-like illness, principally in children (Katagiri, Ohizumi et al. 1987).

Human and animal coronaviruses, *Coronaviridae*

The *Coronaviridae* contains large enveloped positive-stranded RNA viruses, within the order of Nidovirales. The *Coronaviridae* family consists of two genera, Coronavirus and Torovirus (Cavanagh 1997). CoVs causes mild to severe acute upper respiratory infections in humans. In animals, they are often associated with respiratory, gastrointestinal, liver and neurologic diseases. The virion is round, with a diameter of 100-160 nm and distinctive long, petal-shaped spikes on the surface. The genome length of CoVs is rather unique, as they have the largest genome of all RNA viruses, 27-30 kbp (Fauquet and Martelli 1995). In the virion, the genomic RNA is enclosed in the nucleocapsid (N) protein forms the nucleocapsid, which is surrounded by a lipid membrane containing the spike (S), membrane (M), and envelope (E) proteins (Siddell 1995). The 5' end of the viral genome contains two overlapping reading frames (ORFs) called 1a and 1b, which occupies two-thirds of it, ORF 1a and ORF 1b are translated by ribosomal frame-shifting. (Brierley, Digard et al. 1989; Bredenbeek, Pachuk et al. 1990; Lee, Shieh et al. 1991; Herold and Siddell 1993). The gene products of both ORFs are believed to be processed into a number of functional subunits (Denison, Zoltick et al. 1991; Liu, Brierley et al. 1994).

The CoV are divided into three serogroups, based on surface antigens, where serogroup I and II have been isolated from mammals and serogroup III from birds. Avian Infectious bronchitis virus (IBV) was the first CoV isolate (Hudson and Beaudette 1932; Doyle 1946). Classification of the
severe acute respiratory syndrome Coronavirus (SARS-CoV) into group 2 or as the prototype of group 4 is subjected to controversy and is complicated by the putative recombinant origin of its genome.

CoVs generally cause mild respiratory or intestinal infections in mammals and birds, and they are able to generate adaptively useful genotypic variation through recombination in the spike glycoprotein. Two notorious exceptions of CoV cause severe diseases: feline infectious peritonitis virus (FIPV) and acute respiratory syndrome coronavirus (SARS-CoV). FIPV is an important disease that occurs in cats of all ages, a mutant of a harmless virus that causes a highly lethal infection. Human infection by SARS-CoV appears to be limited to the respiratory tract. SARS-CoV is suggested to be a recombinant between different coronaviruses i.e., from both mammalian and avian. (Rest and Mindell 2003; Stanhope, Brown et al. 2004; Stavrinides and Guttman 2004). The emergence of the new SARS-CoV has dramatically changed the clinical relevance of human CoV infections.

**Diagnosis of viral diseases**

In all clinical work the benefits of precise diagnostics are indisputable. The consequences for the treatment of individual patients are obvious, and preventive measures can be taken to reduce the risk of transmitting the infection to others. The diagnosis of viral diseases can in general be based on:

**Direct examination of specimen**
- Electron microscopy that allows detection and identification of virus particles on a morphological basis.
- Light microscopy of histological appearance - e.g. inclusion bodies.
- Antigen detection immunofluorescence, ELISA etc.
- Molecular techniques for the direct detection of viral genomes

**Indirect examination**
- Virus isolation, cultivation and identification of virus grown in tissue culture or eggs inoculated with specimen.
- Cell culture - cytopathic effect, haemadsorption, confirmation by neutralization, interference, immunofluorescence etc.
- Disease in animals or death confirmation.

**Virus serology**
- A serological diagnosis is based on the detection of increasing titres of antibody between acute and recovery stages of infection, or the detection of IgM.
While traditional methods such as virus isolation are still widely used in laboratory diagnostics, it is well recognized that they often are unacceptably slow to provide a report, and require costly facilities and trained personnel. As a result, the diagnostic results arrive too late for the clinician to exert any influence on treatment. Furthermore, cell cultures and viral antigen methods vary greatly in their susceptibility to different viruses and electron microscopy cannot be relied on exclusively as it has poor sensitivity. Therefore, the force of developments in diagnostic of DNA and RNA viral infections have led to molecular assays that detect or quantify viral nucleic acid. The arrival of rapid and sensitive techniques for detecting viral antigens and nucleic acids is now effecting a major improvement in the standard of service that can be offered to the clinician by the routine virology laboratories. Molecular diagnostics has become part of routine clinical laboratory testing for the diagnosis and monitoring treatment of patients infected with RNA viruses such as human immunodeficiency virus (HIV), hepatitis C virus (HCV), enteroviruses, etc.

An advantage of molecular diagnostic technology is that it makes viral load quantification possible, which is of clinical importance. A general rule is that more severe infections have higher number of virus particles. Quantitative determination of viral load using a real-time polymerase chain reaction technique allows follow-up and prognostication of chronic viral infections. It thus adds an important dimension which cannot be provided by serology. In this way, quantitative PCR has revolutionized clinical virology during the last 10 years. It has an unprecedented dynamic detection range, from $10^0$ to $10^8$ target copies per PCR reaction, over which quantification can be made. This interval covers most clinically important questions. The clinical applications of molecular assays help to optimize the patient outcome. Another advantage of molecular diagnostics is characterization of the heterogeneity of viruses.

A number of RNA viruses are highly infectious, acutely pathogenic, and a significant source of morbidity and mortality in humans every year. A common goal in virus research is to measure the amount of viral nucleic acid. Development of rapid and reliable early diagnostic techniques is linked to the success of antiviral therapy for highly pathogenic RNA viral infections, since the infection can be identified at an early stage. Quantification allows the course of the infection, and the degree of therapeutic success, to be followed. Moreover, many viruses cannot be cultivated and some are only known as sequences. Endogenous retroviruses are examples of this. They are mainly studied by nucleic acid based techniques. A RNA virus population can be extremely heterogeneous because of sequence diversity, which makes diagnosis of some virus infections complicated. Real-time PCR techniques often offer a fortuitous combination of rapid, sensitive and
rational detection, three clinically important attributes. Critical clinical situa-
tions may demand high sensitivity, as the low infectious dose of viruses 
requires that methods are able to detect very low numbers of virions in pa-
tient samples. A fact which complicates nucleic acid amplification methods
for RNA viruses is that they require an additional reverse transcription (RT)
step. However, the clinical need is driving technology development forward.
A one-step Real-time PCR may detect single viral nucleic acid molecules.
Aims

Goal and specific aims

The new generation of nucleic acid based molecular diagnostic techniques such as the polymerase chain reaction (PCR) are powerful and sensitive diagnostic tools for identification of viral pathogens even at a very early stage of infection. The general aim of this thesis was to develop broadly targeted real-time PCR methods for detection of variable RNA viruses such as human retroviruses and respiratory viruses, influenza viruses and corona viruses. A variety of diagnostic PCR methods are already available for the study of these viruses. However, in most cases they are narrowly targeted to single or a few viral species.

Specific aims

- To develop and apply new broadly targeted real-time PCR methods for detection and quantification of human betaretrovirus-like sequences.

- To develop and apply a QPCR for DNA of exogenous mouse mammary tumour virus-related virus in human breast cancer samples.

- To develop broadly targeted QPCR methods for the detection of influenza viruses in humans and animals.

- To develop broadly targeted QPCR methods for the detection of corona viruses in humans and animals.
Materials and Methods

Materials

Clinical specimens

**Paper I**
Human brain tissue samples from four patients with multiple sclerosis and four apparently healthy individuals were obtained from the Human Brain and Spinal Fluid Resource center (c/o professor Wallace Tourtellotte, UCLA). A RNA panel from different tissues of healthy people was purchased from Clontech (Human Total RNA Master Panel II, catalog number: K4008-1, Clontech, Palo Alto, Ca, USA).

**Paper II**
Human malignant and nonmalignant tissue and blood was obtained from 18 consecutive patients with palpable breast cancer and 11 control women with a diagnosis of a benign breast condition. Tissues were kept at -70 °C until RNA extraction. The study was approved by the ethical committee of the faculty of medicine of the University of Uppsala (permit number 01/381).

**Paper III**
Nasopharyngeal aspirates from 203 patients with a suspected influenza virus infection during winters 1999-2005, submitted for the purpose of a viral diagnosis, were collected. Eighteen influenza A, two influenza B and five Japanese influenza C isolates, were collected from national and international reference laboratories.

**Paper IV**
During winters 1999-2002, 77 nasopharyngeal aspirates from patients admitted to hospital with acute respiratory syndromes, submitted for the purpose of a viral diagnosis, were collected. RNA from 35 pooled duck fecal samples in total representing 100 wild birds (Ottenby Bird Observatory, Öland, Sweden). A collection of 22 coronavirus reference strains were collected through a number of collaborators, who are acknowledged in the manuscript.
Template handling and preparation

Sample preparation is an important step when performing PCR, and particularly when performing quantitative PCR. In general, PCR is a powerful and rapid method for diagnostic of microbial infections and genetic diseases, as well as for detecting micro-organisms in environmental and food samples. However, despite its potential diagnostic use, PCR inhibitors have been an obstacle to success for researchers, which makes PCR limited since inhibiting substances may reduce or even block the amplification of the template, i.e. giving a false negative result. Likewise, the instability of RNA due to RNAses can lead to RNA loss and false negativity. There are several ways to study the presence of PCR inhibitors or control for RNA loss in a sample such as by amplifying different endogenous control genes (i.e., genes serving fundamental roles in the cell, such as HIS, GAPD, HPRT1 and UBC etc), by spiking the sample with purified DNA or by running a dilution series on the sample. The inhibition of amplification may be due to inhibiting substances that may originate from; ineffective reaction conditions; from the original sample or from preparation prior to PCR or from both; contaminants in reagents, containers, or disposables; or from contamination during reaction preparation. One of the safest approaches to reduce PCR inhibitors is to prevent the inhibitors from being processed with the sample or to use DNA/RNA purification methods that remove inhibitors.
Methods

Development of sequence-based diagnostic tool

The development of all of our nucleic acid-based diagnostic methods was divided into two stages, which include a preclinical- and a clinical stage. Preliminary information on the assay performance was investigated by feasibility (preclinical) analytical studies. Furthermore, all the assays were validated with clinical samples submitted because of a suspected viral infection, handled anonymously.

Preclinical stage

1. Problem definition
   a. Specific viral pathogen detected and identified
   b. Disease associated with infections by the viral pathogen
   c. Epidemiology, prevalence rate and individual at risk for infection

2. Collection of representative sequences for the targeted pathogen
   a. Identification of conserved stretches suitable for primers and probes
   b. Design of primers and probes

3. Trial of test set against real or synthetic sequences
   a. Determination of limits of detection (LoD)
   b. Assay evaluation

Clinical stage

4. Further clinical validation
5. Routine use
Principle of PCR/Real-Time PCR

The polymerase chain reaction (PCR) technique was invented by Kary B Mullis. It derives its name from one of its key components, a DNA polymerase used to amplify a piece of DNA by in vitro enzymatic replication (Mullis, Faloona et al. 1986). As the name suggests, real time PCR is used to monitor the progress of a PCR reaction in real time using fluorescent dyes that intercalate with double-stranded DNA, or modified DNA oligonucleotide probes that fluoresce when hybridized with a complementary DNA. In probe-based systems the key feature is that the amplified DNA is quantified as it accumulates in the reaction in real time after each amplification cycle, when the probe is degraded by the 5´-3´exonuclease activity of the Taq polymerase (Lee, Connell et al. 1993). The SYBRGreen intercalating dye, TaqMan® dual labelled probes and molecular beacon chemistry are the most common monitoring systems.

Probe based Real-time PCR is a technique in molecular biology based on energy shift from an electronically excited molecule (donor i.e., reporter that is a high-energy dye) to neighbouring molecule (acceptor i.e., quencher that is a low-energy dye) through a dipole-dipole coupling mechanism. This energy transfer mechanism between two chromophores is called Förster resonance energy transfer (FRET) (Holland, Abramson et al. 1991; Lakowicz and Masters 2008). The FRET efficiency depends on different parameters such as; the distance between the donor and the acceptor, spectral overlap of the donor emission spectrum and the acceptor absorption spectrum, the relative orientation of the donor emission dipole moment and the acceptor absorption dipole moment.

![Energy level diagram for energy transfer (FRET) from donor to acceptor](image)

**Figure 1.** Energy level diagram for energy transfer (FRET) from donor to acceptor: The electrons of the donor molecule (1) are excited (S₀→S₁) by incoming photons. When the electrons return to the ground state (S₁→S₀) energy is released (2). Because of proximity of the two molecules the energy released from the donor molecule is absorbed by the acceptor molecule and FRET has occurred (S₀). When electrons of the acceptor molecule return to S₀ energy is released at a different wavelength than when it was absorbed (3).
TaqMan® chemistry

TaqMan® probes are dually labelled hydrolysis probes that utilize the 5´-3´ exonuclease activity of the Taq polymerase to measure the amount of generated PCR product during the thermal cycling. The probe is labelled with two fluorophores, a high-energy dye termed reporter at the 5´ and a low-energy molecule termed a quencher at the 3´ end. The 5´ end of the probe is cleaved by the enzyme's double-strand-specific 5´-3´-exonuclease activity and the reporter is physically separated from the quencher. This takes place when the probe is hybridized to its complementary target, resulting in decoupling of the fluorescent and quenching dyes, and FRET no longer occurs i.e., reporter dye’s emission is no longer suppressed by the quencher dye as a result of the close proximity of the dyes (Figure 2). Therefore, reporter fluorescence emission increases in each cycle, proportional to the amount of probe cleaved and displaced (Holland, Abramson et al. 1991; Heid, Stevens et al. 1996). Use of the fluorescent probe in real-time PCR provides an additional level of assay specificity. We used the TaqMan® chemistry in the development of detection and quantification of the HML group of betaretroviruses, MMTV/HMTV, influenza viruses and CoV (Paper I-IV).

**Figure 2.** Representation of how the 5´exonuclease chemistry uses a fluorogenic probe to allow the detection of a specific PCR product. During extension phase the probe is cleaved by the 5´-3´exonuclease activity of the enzyme, the distance between the reporter and the quencher increases causing the transfer of energy to stop. The fluorescent emissions of the reporter can be measured.
Molecular beacon chemistry

Like TaqMan® probes, Molecular beacons (MBs) are single-stranded DNA probes useful for nucleic acid detection, that utilize FRET to detect and quantify PCR product, with a fluorophore attached at the 5’ end and a quencher at the 3’ end. In the absence of target these molecules form a hairpin-like, stem–loop structure which results in fluorescence quenching because of the close proximity of the two fluorophores. During hybridization to the target the fluorophores are apart and hybridization is monitored in real-time by measuring the fluorescence signal. The loop portion is complementary to a predetermined sequence of the target nucleic acid, with a melting temperature ($T_m$) that is 7-10°C higher than the PCR annealing temperature, and the stem has a higher $T_m$ than the PCR annealing temperature (Mhlanga and Malmberg 2001; Bustin 2004; Bustin 2004). In contrast to TaqMan® probes, a molecular beacon remains intact throughout the PCR reaction and is rebound to the target at every cycle (Figure 3). One of the limitations of the method is that the traditional MBs are sensitive to target variation (Broude 2002).

**Figure 3.** Representation of the operation of molecular beacon for detection of PCR product. (A) A molecular beacon remains intact throughout the PCR reaction and is rebound to the target at every cycle. (B) As the PCR continues, PCR products and the molecular beacon are denatured by high temperatures. The hairpin structure is disrupted and FRET no longer occurs. As the temperature cools for the next round of primer annealing, the molecular beacon is capable of forming base pairs with the appropriate strand of the PCR product.
SYBRGreen chemistry

SYBRGreen real-time PCR includes use of a nonspecific dye that binds to double stranded DNA and the SYBRGreen fluorescence is proportional to PCR product accumulated during the amplification (Figure 4). The SYBRGreen method has several advantages, such as the ability to add melting point analysis to the run. The melting curve reveals the degree of heterogeneity of the amplimer, which can signal target sequence changes, allowing judgment of product specificity (Ririe, Rasmussen et al. 1997; Al-Robaiy, Rupf et al. 2001), and allow comparison of the melting temperatures of the specific product and any suspected non-specific products. The limitation is that it detects all double stranded products, including primer-dimers, which may result in an overestimation of the target concentration, and false positive results in cases where the melting curve does not clearly discriminate between primer-dimer and specific product (Lekanne Deprez, Fijnvandraat et al. 2002). A SYBRGreen QRT-PCR was used for housekeeping gene controls in paper I and as a control method in paper IV.

Figure 4. Illustration of how the SYBRGreen dye binds to the double stranded DNA during the extension phase of the PCR.

Reverse transcription polymerase chain reaction (RT-PCR)

Real-time PCR eliminates post-PCR processing of amplification products. This helps to increase throughput and reduce the risks of carryover contamination. Real-time PCR can be combined with reverse transcription PCR, by an additional step of reverse transcription, in which cDNA is synthesized from RNA. The RNA strand is first reverse-transcribed into its DNA com-
plement or complementary DNA (cDNA) by RNA-dependent DNA polymerase (RDDP), followed by amplification of the resulting DNA using polymerase chain reaction. The technique allows quantification of low abundance messenger RNA (mRNA) in gene expression analysis, and is often referred to as quantitative RT-PCR (QRT-PCR), which should not be confused with real-time PCR. QRT-PCR makes it possible to quantify relative gene expression at a specific time, or in a particular cell or tissue type. In fact, this technique is sensitive enough to enable quantitation of RNA from a single cell. Reverse transcription reactions can either be a one or two step process, using target specific primers, random primers or oligo-dT. Oligo-dT primes cDNA synthesis only from the mRNAs that have a poly-A 3' tail and do not prime ribosomal RNA (rRNA) or any other RNA molecules that lacks poly-A 3'. Random primers initiate cDNA synthesis from all RNA species (including rRNA and mRNA) in a total cellular RNA sample. If cDNA from a subset of all mRNA is wanted, then a target sequence specific primer is most efficient to use, which only binds to one mRNA sequence. The use of a target specific primer in reverse transcription is most efficient for cDNA synthesis as compared to priming by random hexamers or oligo-dT, and it allows more sensitive and accurate quantification (Lekanne De Prez, Fijnvandraat et al. 2002), and is used in one-step QRT-PCR (Paper III and IV). However, the choice of an optimal priming strategy may be target dependent. Random primers or oligo-dT or a combination of them is used in two-step QRT-PCR (Paper I and II).

In one-step QRT-PCR it is not possible to store cDNA for later experiments, but it is preferable since it allows quantification in a single reaction which reduces the possibility of cross-contamination and provides a convenient system since it involves less time and effort. The entire newly synthesized cDNA serves as template for the following PCR phase (Goblet, Prost et al. 1989; Wang, Cao et al. 1992; Aatsinki, Lakkakorpi et al. 1994; Mallet, Oriol et al. 1995).

Two-step QRT-PCR systems are generally considered to provide a somewhat higher sensitivity and specificity than one-step QRT-PCR systems. The sensitivity is gained since the reverse transcription reaction and following PCR reaction are run separately, with their own particular optimized conditions, which can differ very much (Bustin 2004).
Broadly targeted QPCR

One of the approaches to achieve broad specificity for a group of heterogeneous sequences is the use of generic, degenerated primers, instead of separate amplification reactions specific for each strain type. Degenerated primers are a set of primers with a number of options at several positions in the sequence, to allow annealing and amplification of a variety of related sequences. A degenerate primer is a primer sequence in which some positions contain more than one possible base by insertion of “wobbles”. The degeneracy of a sequence is the number of unique sequence combinations it contains.

\[ Y = pYrimidines = C / T \text{ (degeneracy = 2X)} \]
\[ R = puRines = A / G \text{ (degeneracy = 2X)} \]
\[ N = \text{Nucleotide} = C / G / A / T \text{ (degeneracy = 4X)} \]

For example, in paper I, we used a primer set with 432- and 144-fold degeneracy as forward and reverse primers, respectively, for amplification of the reverse transcriptase domain in the beta retrovirus group. The forward primer sequence, including the IUPAC ambiguity codes, was as follows: \[ \text{ATTTGCCCTTTACTRDCCWKCHTHAA} \], where \( R=\text{A/G} \) (degeneracy = 2X), \( D=\text{A/G/T} \) (degeneracy = 3X), \( W=\text{C/T} \) (degeneracy = 2X), \( K=\text{G/T} \) (degeneracy = 2X) and \( H=\text{A/C/T} \) (degeneracy = 2X). The sequence above has \( 2^4 \times 3^5 \) variants, which gives a 432-fold degeneracy.

Since only a low fraction of the primer molecules in a degenerate primer pool match the target sequence, amplification efficiency may decrease. One possibility to enhance PCR efficiency is to increase primer concentration, although there will be a greater risk of mispriming and generation of non-specific PCR products. Thus, degeneracy greater than 512-fold should be avoided. However, successful amplification with high PCR efficiency can be achieved with primers with degeneracy as high as 256- and 1024-fold (Forsman, Yun et al. 2005). The forward and reverse primers in the integrase- and reverse transcriptase based PCR in paper I had a degeneracy of 432-, 144-, 432- and 128-fold, respectively.

**Design of degenerated primers**

The major difference in degenerated PCR compared to ordinary PCR is that, instead of using primers with a given sequence, primers with mixed sequences are used. The mixed primers are generated by insertion of “wobbles” where there is more than one possibility. We used to some degree the CODEHOP principle (Rose, Schultz et al. 1998) in methods developed in paper I, III and IV. The principle is to have primers with a two-part structure, with a consensus part at the 5’ end and a 3’ degenerate part that con-
tains all possible nucleotide sequences, The 5´ non-degenerate consensus part contains only the most probable nucleotide, due to a higher mismatch tolerance at the 5´end (Figure 5A-B). If the 3´ end does not base pair with the target sequence, the polymerase will not extend from the primer. Using this strategy primers with a degeneracy of up to 1000 fold (sometimes higher) and still give a good PCR efficiency.

**Figure 5A.** Two-part structure, a consensus part at the 5´ and a 3´ degenerate part where the 5´ end have higher tolerance to mismatch than the 3´ end. B. The perfectly matching variant starts building the amplimers, raising the target concentration which allows less perfectly fitting variants to participate. This is due to the ability of mismatching primers to substitute for each other, giving a concerted action.

**Design of degenerated probes**

Even though “wobbles” can be introduced into TaqMan® probes, probe variants cannot “help” each other like the degenerate primers. Degeneracy at the 5´end should be especially avoided, because the exonuclease activity of the polymerase that starts degradation of the probe at the 5´end can only cleave hybridized nucleotides. In general, mismatches present within the probe-binding site have the potential to destabilize hybridization of the TaqMan® probe resulting in a loss or significant reduction in signal generated. Each degeneration leads to a decreased signal, unless a high $T_m$ allows binding in spite of mismatch. However, a 30 nt probe can generally tolerate 2-3 mismatches. A 25-30 nt distance between fluorophore and quencher cannot be exceeded with good quenching. We therefore used two different approaches to allow hybridization despite higher degree of mismatch. In paper I, we lowered the annealing temperature to 40 °C (i.e., hybridization at low stringency) in order to allow hybridization in spite of mismatch. In paper III, we distributed the degeneracy on three different concomitantly used, relatively short, probes. Locked nucleic acids (LNA) were introduced into two of the probes to
enhance the stability of hybridization to targets that contain nucleotide mismatches. In paper IV we used a novel probe concept (MegB) to design a long fault tolerant probe, which is described in detail in the paragraph; Novel probes.

**Introduction of LNA™ residues to enhance $T_m$**

Locked nucleic acids (LNAs) are synthetic analogues of nucleic acid that contains a modified ribose moiety and are similar to 2'-O-methyl RNA. The ribose moiety of an LNA nucleotide has a bridging methylene carbon connecting the 2' and 4' carbon atoms, that "locks" the ribose ring in the 3' endo structural conformation, which favors RNA A-type helix duplex geometry that is often found in the A-form of DNA or RNA (Kumar, Singh et al. 1998; Petersen, Bondensgaard et al. 2002; Mouritzen, Nielsen et al. 2003; Johnson, Haupt et al. 2004; Ugozzoli, Latorra et al. 2004) (**Figure 6**). The locked ribose conformation enhances base stacking and backbone pre-organization, which results in a rather stable and rigid structure with increased thermal stability (higher $T_m$). Since LNA bases have remarkably high affinity for complementary RNA and DNA sequences they can be combined with natural nucleotides to optimize the hybridization affinity of oligonucleotide probes. Incorporation of one LNA monomer increases the $T_m$ by 1.5 - 3°C (Kumar, Singh et al. 1998; Petersen, Bondensgaard et al. 2002; Mouritzen, Nielsen et al. 2003; Johnson, Haupt et al. 2004; Ugozzoli, Latorra et al. 2004). The increased affinity may facilitate the recognition of nucleic acid targets. The LNA may confer improved mismatch discrimination relative to natural nucleic acids and higher biostability. Their water solubility is high, similar to the solubility of natural DNA or RNA. Because of those properties, modified oligomers can substitute for native nucleic acid for intracellular monitoring applications, disease diagnosis and basic biological studies (You, Moreira et al. 2006).

![Figure 6. LNA monomers contain a modified ribose moiety and are grossly similar to 2'-O-methyl RNA. But the O-methyl group is further constrained in LNA residues and bridges 2' and 4' carbons of the ribose ring. This covalent bridge effectively 'locks' the ribose in the N-type (3' endo) conformation that is dominant in A-form DNA and RNA.](image-url)
Novel probes: Triple probe and MegaBeacon

**Triple-probe concept**

To circumvent the narrow specificity common to probe-based RT-PCRs a triple-probe concept was developed for detection of coronaviruses in human and animal. This detection system is based on a combination of three moderately degenerated overlapping probes i.e. degeneration is distributed on three probes (Probe I, II and III) in order to allow broad detection of CoV from all four CoV groups. In Figure 7, all three probes are aligned with sequences from reference strains of CoV, with one representative from each group.

![Figure 7](image)

**Figure 7.** Alignment of the triple-probes with sequences from reference strains of Coronavirus, one representative from each coronavirus group.

Simultaneous use of more than one probe can be challenging since the involved oligonucleotides must work under the same conditions. For a multiplex reaction to work properly, the probes need to have fairly similar characteristics such as melting temperature ($T_m$) and should not exhibit significant interaction either internally or to one another (probe-to probe interaction). The hybridization strength of an oligonucleotide is determined by its $T_m$, the temperature at which 50% of the probe is hybridized to the target gene. LNA was introduced into probe_I and probe_II in order to achieve approximately same $T_m$ for all three probes. “Wobbles” were introduced to all three probes which gave a degeneracy of 8, 8 and 6, respectively (**Table 3**). These modifications were made to make the probes more resistant to viral variation, allowing hybridization in spite of a few mismatches. Competition between probes did not adversely affect the result.

**Table 3.** Sequence of triple probes and of LNA and “Wobbles”.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Accession No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe_I</td>
<td>5’-[FAM]TTGTATATCAAGAATGGYGTNTTYA(T)G[EDQ]-3’</td>
<td>NC 002645</td>
</tr>
<tr>
<td>Probe_II</td>
<td>5˚-[FAM]TGTTTCTATGTCWGRAWRGCAWAAATGTGG[EDQ]-3’</td>
<td>AY274119</td>
</tr>
<tr>
<td>Probe_III</td>
<td>5˚-[FAM]TCTAATGTGGGGTGDG[EDQ]-3’</td>
<td>NC 005831</td>
</tr>
</tbody>
</table>

**Notes:** LNA positions are underlined and the “wobble” positions are bold
**MegaBeacon probe**

The second probe concept invented is a novel kind of probe called, the "MegaBeacon" (MegB) for detection of influenza A (Figure 8). The increased fault tolerance of MegB allowed detection and quantification of many influenza A subtypes with a single probe. The novel MegaB probe is a fault-tolerant molecular beacon that has TaqMan® properties, i.e. can be degraded by the 5´-3´exonuclease of the thermostable polymerase. Conventional TaqMan® probes cannot be made very long, because the quencher will not be able to quench over long nucleotide distances (>30 nt) (Bustin 2004). The MegB possesses a stem-hairpin structure as seen in conventional molecular beacons. However, in contrast to conventional molecular beacons, a MegB is not only complementary to the target in the loop region but also in the 5´stem (10 nucleotides). It can be over 50 nt long, which gives a high $T_m$ of the target-probe interaction and a high tolerance for target sequence variation. The Taq polymerase enzyme degrades the MegB by its 5´-3´ exonuclease activity and cleaves the dual fluorescent dye-labeled MegB during PCR. On its own, the quencher is bound to the fluorophore probe and MegB does not fluoresce. Three LNAs were introduced at the 5´stem of the probe to improve affinity for complementary DNA sequences (higher $T_m$) and enhance the degradation by Taq polymerase 5´-3´ exonuclease activity, but also to improve stem stability and thereby quenching of unused MegB.

**Figure 8.** Sequence and structure of influenza A MegaBeacon probe, as visualised by Mfold (Zuker 2003). The three first positions at the 5´ stem are LNAs. The red box indicates the 3´ stem that is not complementary to the target (10 nucleotides). The green circle at 5´ indicates reporter (FAM) and the blue circle indicates the quencher (Dabcyl).
Comparison of MegB and conventional MB

Both molecular beacons and MegBs do not fluoresce when they are free in solution. However, when they hybridize to a nucleic acid strand containing a target sequence they undergo a conformational change that enables them to fluoresce brightly. In the absence of targets, below the $T_m$ of the stem, the probe is dark, because the stem places the reporter so close to the nonfluorescent quencher that they can share energy, eliminating the ability of the reporter to fluoresce. FRET occurs. When the probe encounters a target molecule, it forms a probe-target hybrid that is longer and more stable than the stem hybrid. The rigidity and length of the probe-target hybrid prevents the simultaneous existence of the stem hybrid. Consequently, both the MB and MegB undergo a spontaneous conformational reorganization that forces the stem hybrid to dissociate and the fluorophore and the quencher to move away from each other, restoring fluorescence. MBs are designed to remain intact during the amplification reaction, and must rebind to target in every cycle for signal measurement, while MegBs are degraded by the exonuclease activity of the polymerase during the extension of the newly produced PCR-product (Figure 9A). A minor disadvantage of the MegB may be the long length, which may favour formation of competing non-quenching conformations, which in turn lead to weak quenching. Such conformations physically separate the fluorophores from each other and FRET can not longer occur, which could lead to high background signal. However, the background signal can be distinguished from target-specific signal generated from cleaved probes. However, leaky quenching can take place in both cases due to leaky hybridization as a result of the competition between the probe and the complementary target strand, kinetic advantages of the monomolecular stem closure reaction and non-quenching conformations (Figure 9B).

Further PCR optimization

PCR optimization is one of the most important steps during the development of new methods. Beside the optimization strategies mentioned above; generic detection by introduction of “wobbles” into primers and probes, introduction of LNA residues into primers and probes to enhance $T_m$ and novel probe design strategies (triple-probe and MegaBeacon), a number of other PCR parameters were optimized in this work in order to achieve good reproducibility and sustainability of the methods in paper I to IV.
Figure 9A. Fluorescence measurement advantages of a Megabeacon over conventional molecular beacon. The fluorescence signal is indicated by yellow, where the MegB generates higher fluorescence signal when target is present due to fluorescence generated from MegB bound to target and from fluorophores cleaved during extension (higher Δ fluorescence). Minus and plus symbols at the top of the figures denote absence, or presence of target nucleic acid, respectively. Denaturation, probe-target hybridization and extension phases are indicated by three boxes in both cases. Due to the conformation change, FRET no longer occurs because of the long distance between reporter and quencher.

Figure 9B. Possible quenching leakiness of MBs and the longer MegaB probes, due to competing non quenching conformations, kinetic advantages of a mono-molecular reaction and competition with complementary strand of the amplimer. The dotted line symbolizes fluorescence from free fluorophore as described in figure 9A.
Search for representative sequences

The key to successful PCRs lies in the design of appropriate primers. The lack of proofreading during replication causes a great genomic diversity within endogenous retroviruses and RNA viruses (Holland, Spindler et al. 1982). Even the most conserved, functionally constrained regions, can display sequence diversity and affect binding of primers targeted to them (Zanotto, Gibbs et al. 1996). In primer and probe design for detection of RNA viruses it is important to search for sequence conservation in selection of target nucleic acid. We used a computer program called “ConSort”© (J Blomberg et al., unpublished) to identify conserved portions of the viral genomes. Multiple alignment of viral sequences were generated using: freely available program packages CLUSTALW and MegAlign (Thompson, Higgins et al. 1994; Clewley and Arnold 1997) or web-based BLAST and MultAlin programs. The ConSort analysis was made using those alignments. The primers to be used used in a QRT-PCR system were designed to have about the same annealing temperatures, no significant secondary structures and no homo- or heterodimeric complementarities. The oligonucleotide primers did not display significant homology either internally or to one another.

Optimisation of annealing temperature

Gradient PCR was used to determine the optimum annealing temperature. Here, the PCR products were analyzed by electrophoresis (1-2% agarose gels), and the temperature that yielded the strongest band of the right size was chosen. Moreover, all positive specimens were evaluated by electrophoresis on 1-2% agarose gels. Gels were run at 100 volts for one hour, stained with ethidium bromide, and visualized using ultraviolet light. A negative control was included into all PCR runs and was confirmed negative by electrophoresis.

Multiplex QRT-PCR

Multiplex PCR is a variant in which more than one target sequence is amplified using more than one pair of primers. This is a good concept to reduce costs and to improve the diagnostic capacity of the test. We developed a TaqMan® based QRT-PCR for simultaneous detection of influenza A, B and C (Paper III). Fluorescent dyes with different emission spectra were attached to the influenza A, B and C probes, FAM, ROX and Cy5, respectively. Multiplex PCR has a significant diagnostic value, since it requires less sample material than analyzing a sample for different infectious agents separately.


**Touch-down QRT-PCR**

Touchdown (TD) PCR (Don, Cox et al. 1991) is a versatile one-step procedure for optimizing PCRs to obtain specificity even if the degree of primer-template complementarity is not fully known (Hecker and Roux 1996). The success of PCR amplification relies on the specificity with which a primer anneals only to its target (and not non-target) sequences so it is important to optimize this molecule interaction. Whether a primer can anneal only to its perfect complement or also to sequences that have one or more mismatches to the primer, depends critically upon the annealing temperature. Basically, touchdown PCR involves decreasing the annealing temp by 1-2 degrees every second cycle to a 'touchdown' annealing temp. This enhances the specificity of the initial primer-template duplex formation and allows for the enrichment of the correct product over any non-specific product. The primer will anneal at the highest temperature which is least-permissive of non-specific binding that it is able to tolerate. Thus, the first sequence amplified is the one between the regions of greatest primer specificity; it is most likely that this is the sequence of interest. The touch-down procedure was designed in order to favor more specific annealing in the first few cycles. In subsequent cycles, the annealing temperature is decreased to increase the amplification yield that also includes less well-fitting targets. This method is very useful to increase the specificity of PCR, targeting low-copy-number variable targets (Schunck, Kraft et al. 1995; Marchand, Hajdari et al. 2003).

**Amplification efficiency**

Amplification efficiency evaluation is an essential marker in real-time gene quantification procedures. The amount of PCR product is doubled with each cycle, in case of 100 % amplification efficiency. Fluorescence measurement is monitored during each cycle, and the slope in the plot of fluorescence intensity versus cycle number should be equal to the logarithm of the efficiency. The value of the standard curve slope is used to determine the exponential amplification, PCR efficiency, and the regression and correlation coefficients (expressed as $R^2$). Exponential amplification $=10^{-\text{1/slope}}$, and efficiency $=10^{\text{1/slope}}-1$ (Higuchi, Fockler et al. 1993). Amplification efficiency can be affected by numerous factors, as inhibition in the sample, secondary structure interference, the design of primers and probes, and how well the PCR conditions are optimized (Liu and Saint 2002; Tichopad, Dilger et al. 2003). It is of importance to determine the amplification efficiency in both standard samples (synthetic oligonucleotides, plasmids etc.) and unknown samples, since plasmids may amplify better that extracted material, due to high purity and that no other sequences are present for cross-reactions. If the standard curve is generated with a plasmid, and the plasmid is amplified with higher efficiency than the sample targets, it leads to an underestimation of the number of wild-type viral target molecules in the PCR reaction.
Differences in amplification efficiency can lead to under- or overestimation of the number of copies in the sample, due to the relationship between amplification efficiency and the expected target copy number. A 10% reduction in efficiency (2 fold to 1.9 fold per cycle) will result in an 80% drop in the amount of product. An efficiency test with both serial dilutions of sample and of plasmids/synthetic target was performed for all developed methods included in this work. Additionally, a control for PCR inhibition was to run all DNA samples, and two samples only containing water, in the presence of ("spiked with") 10 proviral copies of C3H/HeJ DNA per reaction (Paper II). The amplification efficiency can also be determined by amplifying two different concentration of a sample or a serial dilution of a sample (Tichopad, Dilger et al. 2003).

**Sensitivity and specificity**

The sensitivity and specificity are the properties of the test that tell us about the test accuracy. The *clinical sensitivity* refers to how many infected persons who are correctly (as judged by a trusted reference method) identified by a positive test result. The *analytical sensitivity* discloses also the lower detection limit. The *specificity*, on the other hand, is defined as the number of negative test results for all negative cases. The sensitivity of the QRT-PCRs were determined by titration of synthetic target or RNA from viral strains or plasmids, with 6 observations of each dilution step, into the stochastic zone of 1-10 copies per QRT-PCR reaction. The templates were titrated in 10-fold dilution steps down to complete negativity. The sensitivity test results indicated a sensitivity of 1-10 copies per PCR assay, with a clear stochastic zone of positivity. In order to assess the specificity, the QRT-PCR assays were performed on RNA or cDNA viral nucleic acids from other viruses than the targeted ones. In paper III and IV, the specificity was checked using specimens containing other respiratory viruses than influenza viruses or coronaviruses and no amplified products were detected in any case.

**Stochastic zone in quantitative PCR**

A single target molecule per reaction may be detectable by some PCRs. In order to distinguish one target molecule from the absence of it a number of parallel reactions must be made. Stochastic effects increase the variance of concentration estimates below 10 copies (Morrison, Weis et al. 1998). Since nucleic acid molecules are distinct entities, their quantification have a Poisson distribution (Stenman and Orpana 2001), in analogy with titration of infectious viral particles.
**Contamination precaution**

The extreme sensitivity PCR makes nucleic acid contamination an important source of error. To avoid contamination different strategies were used; aerosol barrier pipette-tips, and performance of master mix preparation, DNA/RNA purification, PCR, and detection in separate rooms. Contamination due to sample-to-sample or run-to-run cross contamination in PCR can result in a false positive PCR outcome. Physically separating the lab areas used for reagent mixing and sample preparation from the area of product analysis is an efficient way of preventing carry-over contamination (Kwok and Higuchi 1989). Furthermore, diluted (10%) bleach solution was used to decontaminate work surfaces, racks and equipments such as pipettes, vortexes and centrifuges before and after use to minimize risk of contamination.
Results and Discussion

Endogenous MMTV-like sequences in human tissues, Paper I

Several endogenous betaretroviruses have been found to be active in animals such as: mouse mammary tumor virus (MMTV) and Jaagsiekte sheep retrovirus (JSRV) (Fanning, Puma et al. 1980; Hynes, Groner et al. 1980; Palmarini, Gray et al. 2001). Recently, betaretroviruslike sequences have been detected in chicken (Jern, Sperber et al. 2005). These viruses were earlier assumed to be largely confined to mammals.

It has been shown that domains in the Gag of endogenous retroviruses in sheep are required during sheep ontogeny (Dunlap, Palmarini et al. 2006; Dunlap, Palmarini et al. 2006) , which correlates with the hypothesis that endogenous retroviruses have benefited the evolution of their host (Boeke and Stoye 1997). In those animals, the endogenous retrovirus Gag interferes with its exogenous form and is believed to protect the host from infection of exogenous retroviruses (Blomberg, Ushameckis et al. 2004).

HERV-K(HML2) polymorphisms can serve as genetic markers for examining human evolution, as they are stable and identical by descent. The ancestral state is known to be the absence of insertion. Many theories on the biological relevance of HERVs have been advanced during the last 20 years. There are few broadly targeted molecular tools for detection and quantification of whole genera of HERVs (Forsman, Yun et al. 2005). In the last decades, there have been several reports on transcriptional activity of HML2. A higher expression of these elements in breast cancer and germ line tumors has been reported, This is not necessary equal with disease causation (Herbst, Sauter et al. 1996; Herbst, Sauter et al. 1997; Herbst, Sauter et al. 1998). Yin et al reported a variation in the expression of HML sequences between individuals (Yin, Medstrand et al. 1997).

Here, we investigated expression of HMLs belonging to the HERV-K family in human tissues, as measured with pol based PCR systems. The devel-
oped quantification strategies was based on conserved regions of reverse transcriptase (RT) and integrase (IN) domains of the pol gene of different human- and animal betaretroviral were used to investigate the level of expression of all HMLs and the HML6 group in human tissues. The expression analyses with both the reverse transcriptase and integrase based PCRs disclosed that the HML groups are more actively transcribed in placenta, testis and brain compared to other tissues (Figure 10A-B). A lower expression was observed by the IN-system. The differences in expression levels by the two systems could be due to random substitutions or deletions in the primer or probe target sequence. Thus, transcripts from some loci may fit better to one of the two systems. A combination of the two systems should give a reasonably unbiased estimation of betaretroviral expression in a tissue.

Taken together, these data show that many HML loci are expressed. Reproductive tissue and brain are major sites of human betaretroviruslike RNA expression. A broadly targeted QPCR for betaretroviruses is a useful tool for further investigation of possible HML involvement in disease, and to validate previous observations of betaretroviruslike HERV in human carcinogenesis (Palmarini, Fan et al. 1997; Blomberg, Ushameckis et al. 2004)
Figure 10 A. Expression of HML sequences in various human tissues by a real-time PCR which broadly targets the reverse transcriptase and integrase regions of betaretrovirus-like sequences. B. HML6 expression pattern in various human tissues, using HML6RT and HML6IN specific primers and probe. The ratios are calculated by dividing the HML RNA expression level (equivalents/2 μg RNA) in human tissues by the expression level of histone 3.3 (copies/μl) in the same tissue.
Exogenous MMTV-like sequences in breast cancer, Paper II

In the last decades a number of laboratories have tried to demonstrate whether an MMTV-like virus contributes to human breast cancer or not. The reports describes a novel human MMTV-like virus (HMTV) in human breast cancer that shares high sequences identity with MMTV (Wang, Go et al. 1998; Etkind, Du et al. 2000; Liu, Wang et al. 2001; Ford, Faedo et al. 2004; Fernandez-Cobo, Melana et al. 2006). However, limited reproducibility of unique molecular observations has lead to a highly controversial debate about the existence of exogenous HMTV and its positive association with human breast cancer (Yin, Medstrand et al. 1997; Witt, Hartmann et al. 2003; Mant and Cason 2004; Mant, Gillett et al. 2004). Zangen et al used several different tools in the search for MMTV-like envelope gene sequences, but did not detect viral sequences in MCF-7 and MB-MDA-231 cell lines and breast cancer tumors (Zangen, Harden et al. 2002). However, Pogo et al (Wang, Holland et al. 1995) claim the discovery of a human retrovirus which is more than 95% identical to MMTV. They found MMTV-like envelope gene sequences in about 40% of human breast cancer tumours and not in normal breast tissues. The Pogo virus thus is suggested to be an exogenous human homologue of MMTV.

In order to investigate the presence of exogenous mouse mammary tumor virus-related virus in human breast cancer samples, we developed a real-time PCR, based on env of MMTV and the published non-endogenous human MMTV-like (HMTV) (Moore, Dixon et al. 1987; Wang, Holland et al. 1995; Wang, Pelisson et al. 2001). The developed method was used to analyze 18 human malignant and nonmalignant tissue and blood from consecutive patients with a palpable breast cancer and 11 control women with a diagnosis of a benign breast condition. None of them had a measurable expression or proviral presence of HMTV.

The PCR was evaluated by amplifying MMTV env DNA from the mouse strain C3H/HeJ, which contains 1 proviral MMTV copy per genome (information from Jackson Laboratories, Bar Harbor, Maine, USA). A dilution series, in six tenfold steps of this DNA, covering from 900 to 0.009 proviral copies per PCR reaction was included in every PCR run. The ability of the PCR to amplify was further evaluated using 10-fold serial dilution of synthetic DNA of HMTV and MMTV (Table 5). Both sensitivity tests revealed a detection limit of 1-10 copies of MMTV target DNA (Figure 11 A-B)
In summary, while MMTV is accepted as the etiological agent of mammary gland neoplasia in mice, the role of betaretrovirus-like viruses, both exo- and endogenous, and the role of retroviral elements in human breast cancer continue to be debated. Our findings correspond to results from other studies that show that HMTV sequences are not detectable in breast cancer samples. For example, a recent study, which used a PCR-based microarray test, could not detect exogenous MMTV-related pol sequences in breast cancer samples (Leib-Mosch and Seifarth 1995). This negative outcome in search for exogenous MMTV-related sequences could be explained either by down-regulation of transcription at the time of sample preparation or absence of these sequences in the examined tissue.

Table 5. Synthetic HMTV and MMTV oligonucleotides with a size of 140 bp used as positive controls to verify the sensitivity.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synth_MMTV</td>
<td>5'-TAGTCGCTACAAGATATTGTTCTGCTAGCTACTTCTTAATATTCCTCTCTCCGTGTTTTTGCTGCT</td>
<td>M15122</td>
</tr>
<tr>
<td></td>
<td>TAGTTGCAGCCTCAAGATCAAGAGGCCGCCAATATGTGCTGCTACCTGTAGATATTGGTGATGA-3´</td>
<td></td>
</tr>
<tr>
<td>Synth_HMTV_1</td>
<td>5'-TAGTTGCTACAAGATATTGTTCTGCTAGCCTACTTCTTAATATTCCTCTCTCCGTGTTTTTGCTGCT</td>
<td>AF346816</td>
</tr>
<tr>
<td></td>
<td>CTCCTTGTAGATCTTTTCTCTAGCTAGCTACTATGCAACAGTATGTCATAGTCAGAG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCCGCCCTACGTGCTACCTGTAGATATGTTGGTTAATA-3´</td>
<td></td>
</tr>
<tr>
<td>Synth_HMTV_2</td>
<td>5'-TAGTTGCTACAAGATATTGTTCTGCTAGCCTACTTCTTAATATTCCTCTCTCCGTGTTTTTGCTGCT</td>
<td>AF239172</td>
</tr>
<tr>
<td></td>
<td>AATTGTTGATCTTTTCTCTAGCTAGCTACTATGCAACAGTATGTCATAGTCTTTGAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCCGCCCTACGTGCTACCTGTAGATATGTTGGTTAATA-3´</td>
<td></td>
</tr>
</tbody>
</table>

* Positions of primers and probe are underlined. Mismatches versus the probe are bold and italicized.

Figure 11A. The curves represent, from left to right, 900 to 0.009 MMTV proviral copies per PCR reaction. B. From left to right, 10^4-10^9 copies of synthetic HMTV per PCR reaction. No signals were obtained in negative control.
Simultaneous detection of influenza types A, B and C, Paper III

Influenza types A and B are responsible for epidemics of respiratory illness ("flu") that occur almost every winter and are major causes of morbidity and mortality worldwide. All over the world approximately, 100 millions and from 0.5 million to 1 million patients die annually due to influenza infection or post-influenza complications that are the cause of death after influenza infection. Although not much studied, Influenza C seems to differ from A and B types in some important ways. It has never been connected with a large epidemic, and may just cause mild respiratory infections similar to the common cold. Thus, influenza type C infections may not have the severe public health impact of influenza types A and B.

Typical symptoms of influenza illness can include any or all of these symptoms: fever, muscle aches, headache, lack of energy, dry cough, sore throat, and runny nose. The flu illness caused by the influenza virus tends to be worse than illnesses caused by other infectious agents which cause a 'flu-like' illness, including, but not limited to, Mycoplasma pneumoniae, adenovirus, respiratory syncytial virus, rhinovirus, parainfluenza viruses, and Legionella spp. Symptoms of fever and body ache can last for 3-5 days, and then usually gradually ease over several days. Irritating cough and lack of energy may persist for a week or more after other symptoms have gone. Most people recover completely within 1-2 weeks. Influenza is highly contagious and spreads predominantly by droplets when people cough and sneeze and by indirect contact from respiratory secretions getting onto hands, tissues, etc. The incubation time for influenza (the number of days from when you are exposed to the virus to when symptoms develop) ranges from 1 to 5 days, but the average is 2 days. However, as the signs and symptoms of influenza are not specific and often resemble those caused by other infectious agents, diagnosis based on clinical symptoms can be challenging. It is important to have a positive influenza diagnosis in the first cases during outbreaks of respiratory illness.

Influenza virus nucleoprotein (NP), a core antigen of influenza virus, is the basis for dividing influenza strains into type A, B and C. It is a more conserved protein than the membrane glycoproteins (Epstein, Kong et al. 2005), and is therefore suitable for design of a broadly targeted QRT-PCR. The broadly targeted triplex test, for influenza A, B and C, was based on TaqMan® probes targeting influenza B and C viruses (3QRT-PCR) and a further developed assay with novel kind of probe, MegB, for influenza A (3QRT-PCR-MegB).
The influenza triplex 3QPCR was evaluated by analysing 203 nasopharyngeal aspirates from patients with a suspected influenza virus infection, and 18 influenza A, two influenza B and five Japanese influenza C isolates. The real time PCR assays were able to detect 1-10 viral and synthetic cDNA copies per PCR reaction for influenza A, B and C. Moreover, 203 human nasopharyngeal aspirate samples from individuals with respiratory infection were analyzed with 3QPCR. Of the 203, 72 were influenza A, 39 were influenza B and 1 influenza C, consistent with antigen detection by immunofluorescence and a nested influenza A and B PCR. There was no alternative influenza C detection method. 71 of the 203 samples were analysed with 3QPCR-MegB; 28 came out as influenza A, 10 as influenza B. Three of the influenza A samples positive by 3QPCR-MegB were missed by 3QPCR and IFA.

The MegB probe had a high fault tolerance and allowed detection and quantification of 18 influenza A HxNx subtypes. The broadness of MegB was determined by generating a multiple alignment of influenza A strains of different wild type variants representing 25 HN combinations to identify the proportion of variants existing in the probe region (Table 6). The alignment was then used to design synthetic oligonucleotides, with perfect match and up to 10 mismatches in the probe region to cover all variants in the alignment. These oligonucleotides were tested by the 3QRT-PCR-MegB. All of them were detectable.

### Table 6. Matching and mismatching Influenza A synthetic oligonucleotides.

<table>
<thead>
<tr>
<th>Influenza A</th>
<th>Sequence 5’-3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect_Match</td>
<td>GAAATCATAAGAATGATGGAAAGTGCCAGATTCGAGCTCTCGGACGAAAAGGCAACGAACCCGATCGTGCCTTCCTTTGACATGAGTAATCTTATTTCTTCGGAGACAAT</td>
</tr>
<tr>
<td>1_Mismatch</td>
<td>GAAATCATAAGAATGATGGAAAGTGCCAGATTTGAGCTCTCGGACGAAAAGGCAACGAACCCGATCGTGCCTTCCTTTGACATGAGTAATCTTATTTCTTCGGAGACAAT</td>
</tr>
<tr>
<td>2_Mismatch</td>
<td>GAAATCATAAGAATGATGGAAAGTGCCAGATTTGAGATTTCGGACGAAAAGGCAACGAACCCGATCGTGCCTTCCTTTGACATGAGTAATCTTATTTCTTCGGAGACAAT</td>
</tr>
<tr>
<td>3_Mismatch</td>
<td>GAAATCATAAGAATGATGGAAAGTGCCAGATTTGAGATTTCAGACGAAAAGGCAACGAACCCGATCGTGCCTTCCTTTGACATGAGTAATCTTATTTCTTCGGAGACAAT</td>
</tr>
<tr>
<td>4_Mismatch</td>
<td>GAAATCATAAGAATGATGGAAAGTGCCAGATTTGAGATTTCAGACGAAAGGGCAACGAACCCGATCGTGCCTTCCTTTGACATGAGTAATCTTATTTCTTCGGAGACAAT</td>
</tr>
<tr>
<td>5_Mismatch</td>
<td>GAAATCATAAGAATGATGGAAAGTGCCAGATTTGAGATTTCAGACGAAAAGGGCCACGAACCCGATCGTGCCTTCCTTTGACATGAGTAATCTTATTTCTTCGGAGACAAT</td>
</tr>
<tr>
<td>6_Mismatch</td>
<td>GAAATCATAAGAATGATGGAAAGTGCCAGATTTGAGATTTCAGACGAAAAGGGCCGCGAACCCGATCGTGCCCTCCTTTGACATGAGTAATCTTATTTCTTCGGAGACAAT</td>
</tr>
<tr>
<td>7_Mismatch</td>
<td>GAAATCATAAGAATGATGGAAAGTGCCAGATTTGAGATTTCAGACGAAAGGGCCGCGAACCCGATCGTGCCCTCTTTTGACATGAGTAATCTTATTTCTTCGGAGACAAT</td>
</tr>
<tr>
<td>8_Mismatch</td>
<td>GAAATCATAAGAATGATGGAAAGTGCCAGATTTGAGATTTCAGACGAAAGGGCCGCGAACCCGATCGTGCCCTCTTTTGACTTGAGTAATCTTATTTCTTCGGAGACAAT</td>
</tr>
<tr>
<td>9_Mismatch</td>
<td>GAAATCATAAGAATGATGGAAAGTGCCAGATTTGAGATTTCAGACGAAAGGGCCGCGAACCCGATCGTGCCCTCTTTTGACTTGAGTAATCTTATTTCTTCGGAGACAAT</td>
</tr>
<tr>
<td>10_Mismatch</td>
<td>GAAATCATAAGAATGATGGAAAGTGCCAGATTTGAGATTTCAGACGAAAGGGCCGCGAACCCGATCGTGCCCTCTTTTGACTTGAGTAATCTTATTTCTTCGGAGACAAT</td>
</tr>
</tbody>
</table>

Notes: A new Influenza A multiple alignment was generated: in an assessment of 2500 Influenza A NP sequences from GenBank (2008), the 10 displayed variant positions were found. Thus, the MegB target region contained up to 10 mismatching nucleotides versus the MegB probe. Oligonucleotides of all 10 variants were purchased from Thermo Hybaid to test the fault tolerance of the MegB.

To prevent large-scale epidemics and outbreaks of respiratory infections associated with influenza viruses it is important to have rapid diagnosis, early therapeutic interventions with costly antiviral drugs are being considered, rapid testing for influenza virus plays a role in hospital infection con-
trol in reducing the spread of infection from patient to patient or from infected health care workers to high risk patients. The 3QPCR-MegB influenza triplex assay described in our study offers an accurate, sensitive and rapid detection of influenza viruses. The MegB generic nature should enable detection not only of most human Influenza A viruses, but also of avian and other animal influenza strains. The MegB probes hold great promise for detection and quantification of variable vial target nucleic acids.

Detection of coronaviruses in humans and animals, Paper IV

The aim of this study was to develop broadly targeted QPCR methods for the detection of corona viruses in humans and animals. We used a pair of degenerate primers allowing amplification of a 179 nt conserved stretch of open reading frame 1b, which encodes replicase. We present the result of screening 77 nasopharyngeal aspirate samples from human respiratory infections in Uppsala, Sweden (1999-2002) for corona virus. Eight (9%) were CoV positive. Phylogenetic analysis of the sequenced amplicons revealed six of them to be OC43-like and one of them to be NL63-like (Figure 12). Sequences from three samples were not complete, and were not included in the phylogenetic neighbor-joining tree. However, the three, incompletely sequenced samples, yielded similar sequences as the ones belonging to group 2.

The method was used on human and avian samples. In the latter, an unexpectedly high frequency of avian corona viral strains were detected, illustrating the usefulness of the method for epidemiological studies in different host species. The method proved to be more sensitive than our previously published pan-CoV PCR (Escutenaire, Mohamed et al. 2007), because it picked up seven positive samples which were not found by the previous method. Our finding confirms and extends a previous report of CoV in ducks (Jonassen, Kofstad et al. 2005).

Analysis of coronaviral sequences obtained from duck samples revealed that seven of 35 pools of duck faecal samples of wild birds were CoV positive, (100 mallard duck fecal samples of from Ottenby, Sweden and pooled in 11 autumn 2003 and 24 pools spring 2004) (Figure 12). To confirm the amplified products originated from corona-like viruses, the amplicons were sequenced, and along with representative sequences from known CoV family members obtained from GenBank, were used to construct phylogenetic trees. The detected viral strains all were similar to the avian CoV Infectious Bronchitis Virus (IBV) and a pheasant CoV (Figure 12). This indicates that ducks
are not only reservoirs of influenza A virus but also that they rather often have CoV in their gut, and may play a role in CoV epidemiology.

**In conclusion**, QRT-PCR methods are safer and faster than current cell-culture based assays for CoV detection (Huang, Lin et al. 2005). Our triple-probe pan-coronavirus QRT-PCR is valuable for generic detection of coronaviruses, including uncharacterised variants of human or animal. The assay identified eight CoV positive samples of 77 human nasopharyngeal aspirates. Additionally, seven of 35 avian pool samples which were negative with an earlier pan-CoV QPCR came out positive with the triple-probe QRT-PCR.

---

**Figure 12.** Phylogenetic, neighbor-joining, tree. based on the short (179bp) amplicon stretch. The four major coronavirus groups are shown as boxes. Eight CoV positive human samples were sequenced, and aligned with reference sequences. Four of the seven sequenced amplicons turned out to belong to CoV group 2, being OC43/MHV-like (Human Uppsala 2002, Human Uppsala 1958, Human Uppsala 1903 and Human Uppsala 2166). Six of the seven sequenced amplicons from mallard ducks (Mallard_Ottenby_Autumn2, 3, 4, 5 and 11) and 2/11 (Mallard_Ottenby_Spring 3 and 13) resembled the peafowl CoV and Avian infectious bronchitis virus. The sequence from one sample was not complete, and was therefore excluded from the tree (Mallard_Ottenby_Autumn 11).
Concluding remarks

The studies included in this thesis are a collection of developed fluorescence PCR detection methods and applications that can be used for epidemiological studies and diagnostics of RNA- and DNA virus infections. Experience and information from each individual study was gathered and included in this thesis including the benefits, disadvantages and challenges.

The human genome contains a variety of MMTV-like sequences, the HMLs. These are dispersed over different chromosomes and grouped into HML groups 1-10 (HML1-10). The human genome contains around 800 HML proviral copies. Despite many efforts, it is not established whether betaretrovirus-like viruses, exo- or endogenous, are involved in the etiology of breast cancer, or other cancer diseases, in humans. We therefore developed a general betaretroviral QRT-PCR and a HMTV specific PCR that is useful for further investigations on the pathophysiological contribution(s) of human betaretrovirus-like sequences. Despite a high sensitivity of the HMTV detection QPCR, it came out negative in the Uppsala breast cancer samples.

CoV have come into focus as widespread more or less severe pathogens. They are widespread among vertebrates, and recombination between coronavirus strains seems common. Previous broadly targeted Coronavirus detection systems are either not generic enough, or are SYBRGreen based. The latter technique lends itself to a broad detection range, but has the disadvantage that weak specific signals can be obscured by primer-dimer signals. We therefore designed a pan-CoV QPCR which utilizes the TaqMan® principle. However, the necessary degeneracy of a single broadly targeted TaqMan® probe for the entire CoV family is prohibiting, with 768 variants and gave very low signal strength. We therefore worked out a new approach, to distribute the degeneracy on three different, in tandem used, probes. Each of them had degeneracy below 10. The new strategy proved to work. Signal strength was acceptable, and the detection range encompassed all CoV groups.

Finally, we also invented a new probe concept, MegB, to overcome the problem of false negative results with variable viruses. MegB have a high mismatch tolerance because the hybridisation is distributed over a long stretch. They allow broad targeting and detection of highly variable targets.
The novel probe strategy detection system is rapid, specific, and sensitive. The new probe construct works well together with TaqMan® probes for detection of influenza B and influenza C virus in a triplex system and should be useful for routine diagnostic testing.

The worldwide severe acute respiratory syndrome (SARS) outbreak in 2002-2003 caused by a novel corona virus claimed over 800 lives and affected more than 30 countries (Drosten, Gunther et al. 2003; Peiris, Yuen et al. 2003; Satija and Lal 2007). The SARS and Avian Influenza outbreaks are examples which show that broadly targeted QPCRs are important for detection of emerging infections arising from new viral strains. Identification of viral infections depends on both laboratory detection and surveillance intensity.
Acknowledgments

The best and worst moments of my doctoral dissertation journey have been shared with many people. The completion of my thesis would not have been possible without the help of many people who gave their support in different ways. To these people I would like to express my gratitude and sincere appreciation. I am truly and deeply indebted to so many people that there is no way to acknowledge them all, or even any of them properly. I sincerely hope that everyone who knows that they have influenced me feels satisfaction that they have helped such a poor soul along and does not feel remorse that either I did not get it right or then ungratefully omitted them from explicit mention.

First of all thanks to my supervisor professor Jonas Blomberg for giving me the opportunity to do my PhD thesis at his lab. As research never goes smoothly it is important to have someone believing in you and I felt that way all through the years. You have been an inspiring mentor, challenging my ability to conduct research.

My co-supervisor Göran Friman, for his contribution to my PhD-program.

My appreciation goes to all of the collaborators; Prof. Sándor Belák, Alia Yacoub and Akos Hornyak at the Department of Virology at National Veterinary Institute. Prof. Björn Olsen at the section of Infectious Diseases. Associate Prof. Björn Herrman of clinical Microbiology, and Jan Fohlman at Department of Infectious Diseases at Växjö University.

Associate Prof. Bengt Kalin at the routine laboratory, for shearing your long experience in molecular biology.

My appreciation goes to Prof. Hans Nordgren at the Department of Genetics and Pathology, Dr. Fredrik Wärnberg, and associate professors Ewa Lundgren and Anders Liss at the Department of Surgery, for valuable help with collecting breast cancer tissue samples.

All my co-workers at Clinical Virology; Guma Abdeldaim and Farid Benachenhou, for lively discussions about religion and middle-east politics. Christina Öhrmalm, for being a wonderful friend, always paying attention to
my worries, giving me advice, whether it has been work-related or private. Ronnie Eriksson and Magnus Jobs, for making workplace a more pleasant environment with your positive attitude.

My gratitude goes to previous co-workers. Anna Forsman and Zhihong Yun, for introducing me to the world of PCR and cloning. Amarinder Bindra, for introducing me to my second project and shearing your great knowledge in medicine. Nahla Mohammad, Amal Elfaitouri, Lijuan Hu and Robinah Kisekka, for your positive attitude and for being great office mates.

The researchers at section of Bacteriology; Eva Hjelm, Ylva Molin, Markus Klint, Marie Edvinsson, Petra Edqvist, Karin Grannas, Kristofer Severinsson, Ioana Bujila and Linus Kristersson, for creating a delightful atmosphere during lunch and coffee times. Special thanks to Markus for helping me out with the sequencing.

The Thursday’s virus-meetings people; Kåre Bondesson, Johan Lennerstrand, Anders Bergqvist, Bosse Albinsson and Malin Svensson, for great virology and molecular biology meetings.

The hygiene nurses and medical laboratory technicians; Margareta Edvall, Camilla Artinger, Eva Tano, Bawantee Dudhee and Jamila Mohammad at Public health (Vårdhygien), for all the nice talks and support. Margareta, I will miss hearing you give advises on the phone. Eva, for your support and encouragement.

The medical laboratory technicians at the routine laboratory. Sultan Golbob, for being a great officemate for a short time, and for all the annoying yet charming mockery about us Kurds and our lack of independent country. Christina Vincentsson, for helping me with the sequencing. Kamerun Nahar and Ylva Persson, for always being so kind when I have asked to borrow the PCR machines at the routine.

The lovely Iranian ladies in the building; Nahle Shaverdy, Mehrangiz Moghadam, Parisa Haraty and Saideh Berenjian, for the all the wonderful laughter during Fika times.

The administrative personnel; Eva Haxton, Birgitta Sembrant, Helena Ström and Ulla-Britt Stolt, for giving me administrative help during these past five years and always with a smile.

Anders Malmsten. For spending a lot of time helping me to make the lay-out of the thesis, giving me enlightening comments on the text and figures, which allowed me to complete the thesis on schedule.
Prof. Amanda Clinton, for carefully reading and commenting this thesis.

My best friends Azadi Taghavi and Jila Ashaier and Ann-Catrin Olsson “Tin-Tin”. Thank you for always given me positive energy and for encouraging me, in your own individual ways, to always strive to be a better person. Azadi, thank you for all of those times you told me, “Shmu you will make it”. Jila, for your all the nice after work drinks and for putting up with me all those times I have been late or cancelled our meetings because of lab-work. Tin-Tin, for all the long phone talks.

My dear friend Asos Shafeek. For introducing me to the world of charity-work and giving my other things than research to think of, for all computer help. Your amazing spirit and unlimited engagement in everything you do have been great inspiration to me. Thank you!

My wonderful eight sisters, loving brother, mother and her husband, thank you for believing in me and always have been supportive no matter what I have done in my life. Each and one of you have a special place in my heart. I am fortunate to have you as family who consistently and selflessly give me your love, support, and encouragement. Love you all!

Of course, my husband Fuad Bahram, for all your support and unconditionally love during my good and bad times. A lifetime with you will always be too short! My son, my heart, Lawand Bahram, for reminding me every day that we must not forget our spirit and our passion, because that is what makes everything in life worthwhile. I thank you for being patient, for sharing me with my work, and for bringing a smile to my heart.

Shaman
References


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 365

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine”.)