CHARACTERIZATION OF VIRAL PROTEASES FROM NORWALK VIRUS, POLIOVIRUS, AND TRANSMISSIBLE GASTROENTERITIS VIRUS USING A FLUORESCENCE RESONANCE ENERGY TRANSFER ASSAY

by

VENKATA KIRAN PASUPULLETI

B.S., ANDHRA LOYOLA COLLEGE, 2009

A THESIS

Submitted in partial fulfilment of the requirements for the degree

MASTER OF SCIENCE

Department of Diagnostic Medicine/Pathobiology
College of Veterinary medicine

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2012

Approved by:

Major Professor
KYEONG-OK CHANG

Abstract

Positive sense RNA viruses include diverse groups of viruses that cause a wide variety of diseases in humans and animals. Most of these viruses encode proteases that cleave the viral polyprotein into intermediate or mature functional proteins during virus replication. As these proteases play a critical role in virus replication, they represent an attractive target for the development of antiviral drugs. In this study, the main goal was to establish assay systems and characterize the enzymatic activity of related proteases from Norwalk virus (NV), poliovirus, and transmissible gastroenteritis virus (TGEV). These proteases share several common characteristics including a typical chymotrypsin-like fold, a Cys residue as a nucleophile in the catalytic triad (or dyad) composed of Cys, His and Glu (or Asp) residues, and a preference for a Glu or Gln residue at the P1 position on the substrate. We cloned and expressed proteases from these viruses and characterized their enzymatic activities using a fluorescence resonance energy transfer (FRET) assay using a specific FRET substrate corresponding to each viral protease. First, assay conditions of the FRET assay was optimized for each virus protease. Second, inhibition profiles of each virus protein were investigated using five commercially available standard protease inhibitors (chymostatin, leupeptin, antipain, TPCK, and TLCK). The inhibition studies showed that TPCK inhibited NV, poliovirus, and TGEV proteases with varying strength, and chymostatin inhibited only NV protease. All other inhibitors had little effects on the virus proteases. The established FRET assays should facilitate screening potential antivirals.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my major professor, Dr. Kyeong Ok Chang. Without his wisdom and expertise, motivational skills and probably most importantly, his patience, I would never accomplished anything close to this. Thanks to my committee members, Dr. Om Prakash and Dr. Oberst R.D for their expertise. Thanks to David George for all of his technical advice and expertise. Special thanks to Dr. Kim for her continuous support. Thanks to all of the members of Dr. Chang lab and other members of the KSU CVM DMP for all their help.

CHAPTER 1- PROTEASES LITERATURE REVIEW

Introduction:

A large number of biochemical reactions that occur in life are almost mediated by a remarkable array of biological catalysts known as enzymes. Biochemical reactions that are catalyzed by the enzymes are involved in all facets of cellular life: metabolism (the chemical processes that occur within a living organism in order to maintain life by producing cellular building blocks and energy from food sources); biosynthesis (the production of complex molecules within living organisms or cells); detoxification (the process of removing toxic substances or qualities from cells); and information storage (the process of deoxyribonucleic acid synthesis (18) . Proteases also termed as peptidases or proteinases are enzymes which are widely distributed in cells. They catalyse the hydrolysis of peptide bonds that link amino acids together in the polypeptide chain forming the protein. This process is called proteolysis **(Figure 1.1)**. Proteases are present in plant and animal tissues and also in microorganisms. About 2% of the human genome consists of proteases, whereas in case of infectious organisms and other form of life, they account for about 1-5% of the genomes (43), (1). Proteases form one of the extensive and most important groups of enzymes and involved in many important physiological processes occuring in living organisms like protein turnover, digestion, blood coagulation and wound healing, fertilization, cell differentiation and growth, cell signalling, immune response and apoptosis (96).

Proteases are found in different forms either soluble or associated with plasma membranes and subcellular organelles, or even secreted (7). They are found intracellular and their activity has been reported in many tissues, but most of their activity was intensively studied in the brush border branches of the intestine and kidney (57), (101), the anterior pituitary (71), leucocytes (61), erythrocytes (95) and spermatozoids (129) in vertebrates.

Plant genomes encode hundreds of proteases which are key regulators of a striking variety of biological functions in plants including meiosis, gametophyte survival, embryogenesis, seed coat formation, etc.(135). Proteases are also found in microbes like bacteria, fungi and viruses with broad biochemical activity and susceptibility to genetic manipulation (116). Microbial proteases are preferred than plant and animal proteases, because they possess almost all the characteristics required for the biotechnological applications (102). This very high extensive presence of proteases in both organs and subcellular compartments suggests that they play an important biological role in both physiological and pathophysiological conditions of an organism (87).

Sources of proteases:

Proteases are ubiquitous, because they are physiologically important for all living organisms. They are found in a wide diversity of sources such as plants, animals and microorganisms (102).

Plant proteases: Some of the well-known proteases of plant origin are papain, bromelain, keratinases, and ficin. Papain is a traditional plant protease which is extracted from the latex of *Carica papaya* fruits (120). Papain is extensively used in industry for the preparation of highly soluble and flavoured protein hydrolysates. Bromelain is obtained from the stem and juice of pineapples. Some group of plants produces proteases which degrade hair (102).

Animal proteases: The well-known proteases of animal origin are pancreatic trypsin, chymotrypsin, pepsin, and rennins (13), (47). Trypsin is a serine protease which is an important intestinal digestive enzyme responsible for the hydrolysis of food proteins. It is also used in the preparation of cell culture media and in some medical applications. Chymotrypsin is found in animal pancreatic extract which is activated by trypsin from its inactive form called chymotrypsinogen. Chymotrypsin is specific for the hydrolysis of

peptide bonds in which the carboxyl groups are provided by one of the three aromatic amino acids. Pepsin is an acidic protease present in the stomach. It catalyzes the hydrolysis of peptide bonds between two hydrophobic amino acids (102).

Microbial proteases: proteases can be produced extracellularly and intracellularly by microorganisms. Bacteria, fungi and viruses are capable of producing proteases. Microbial proteases have more commercial applications than proteases of plant and animal origin. This is because of the low yield and difficulty in the production of proteases from plant and animal sources. On the other hand the process of protease isolation from animal and plant sources requires a series of extraction and purification steps (102). Bacteria and fungi produces neutral and alkaline proteases, additionally fungi produces acid proteases. Bacterial and fungal proteases are important to the global carbon and nitrogen cycles in the recycling of proteins (32). A secreted bacterial protease may also act as an exotoxin, and be an example of a virulence factor in bacterial pathogenesis (81). Viral proteases have gained importance because of their functional involvement in the processing of proteins of viruses that cause certain fatal diseases in animals and humans. Serine, aspartic and cysteine proteases are found in various viruses (105). Viral proteases are required for the processing of viral polyproteins or the maturational processing of precapsids, and their catalytic activity is required for the production of new, infectious virion (132). Now a days an extensive research has focussed on the three dimensional structure of viral proteases and their interaction with synthetic inhibitors to design potential antiviral drugs.

Classification and nomenclature:

A milestone in the development of any field of study is the existing of a sound system of nomenclature and classification for the objects with which it deals. Both

nomenclature and classification are essential for information-handling. A better system also serves to emphasize important questions and thus provokes new discoveries.

Proteases are classified in subgroup 4 of group 3 (hydrolases) by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (153). However general system of enzyme nomenclature does not comply easily for proteases, because of their huge diversity of action and structure. Based on three major criteria, currently proteases are classified: (i) type of chemical reaction catalyzed by the protease, (ii) chemical nature of the catalytic site, and (iii) evolutionary relationship with reference to the structure (108), (107). Proteases are subdivided into two major groups, i.e., exopeptidases and endopeptidases, based on their site of action. Exopeptidases cleave the peptide bond proximal to the amino or carboxy termini of the substrate, whereas endopeptidases cleave peptide bonds distant from the termini of the substrate (6). Proteases are further classified into four important groups, depending on the functional group present at the active site of the enzyme: (i) serine proteases, (ii) cysteine proteases, (iii) aspartic acid proteases, and (iv) metalloproteases (40). Apart from these four groups, there are two small groups of proteases namely threonine and glutamic acid proteases which were not described until 1995 and 2004, respectively. The other four groups of proteases have been intensively studied for both their catalytic mechanisms and proteolytic activity (40), (44), (109), and (110). There are few proteases which do not precisely fit into the standard classification, e.g., ATP-dependent proteases which require ATP for their catalytic activity (79). Proteases are classified into different families, depending on their amino acid sequences (4) and further subdivided into "clans" to get suitable sets of peptidases that have diverged from a common ancestor (105).

Classification of proteases by IUBMD (based on the details of the reaction catalyzed)

The term “peptidase” has been suggested by the International Union of Biochemistry and Molecular Biology (IUBMB) to represent a subset of peptide bond hydrolyses with code E.C.3.4. “Protease” is the word widely used as synonym to peptidase. In 1992 IUBMB classified peptidases into two categories: (i) Exopeptidases, and (ii) Endopeptidases (153). The IUBMB peptidase/protease classification is shown in the **Table 1.1**

Exopeptidases:

They are classified according to their site of action. They act only near the ends of polypeptide chains. Based on their site of action at the N or C terminus, they are classified as amino and carboxypeptidases, respectively (6). Aminopeptidases act at a free N terminus of the polypeptide chain and liberate a single amino acid residue, a dipeptide, or a tripeptide. They are present in a wide diversity of microbial species including bacteria and fungi (114). Most of the aminopeptidases are intracellular enzymes, but there is a report on an extracellular aminopeptidase produced by *Aspergillus oryzae* (67). Carboxypeptidases act at the carboxyl terminal ends of proteins. Based on the nature of amino acid residues at the active site of the enzymes, carboxypeptidases are dived into three major groups, serine carboxypeptidases, cysteine carboxypeptidases, and metallo carboxypeptidases (102).

Endopeptidases:

Endopeptidases act on the peptide bonds in the inner regions of the polypeptide chain away from the N and C termini (6). The endopeptidases, which cleave peptide bonds at points within the protein, are often called as proteinases in the literature (153). These enzymes show negative activity in the presence of free amino and carboxyl groups. On the basis of their catalytic mechanism, endopeptidases are divided into four sub groups, (i) serine proteases, (ii) aspartic proteases, (iii) cysteine proteases, and (iv) metalloproteases (40).

Based on this system of classification, every peptidase is given a single four-number code. The codes of proteases that will be discussed on in this dissertation are: Norovirus 3C-like protease (E.C 3.4.22.66), Poliovirus 3C protease (E.C 3.4.22.28), and Porcine transmissible gastroenteritis virus (E.C 3.4.22.69). This classification system provides a practical approach for the characterization of new peptidases, without requiring its structure information. Enzyme nomenclature has been used for many decades and it depends mainly on the reaction catalyzed, but failed to recognize the structural relationship among peptidases (153).

Classification of proteases based on the catalytic or chemical mechanism

In 1960, Hartley initiated a series of developments in the protease community that has now provided the very useful idea of catalytic type. Based on the mechanisms used in the catalysis, proteases are currently classified into six groups namely serine proteases, cysteine proteases, aspartic proteases, metalloproteases, threonine proteases, and glutamic acid proteases. These proteases are named based on the active amino acid in the catalytic site of the protease. For example, cysteine proteases contain a cysteine amino acid residue in the active site that acts as nucleophile in the catalytic site of the protease. Among the six groups of proteases, only four groups of proteases (serine, cysteine, aspartic, and metalloproteases) are extensively studied for both their catalytic mechanisms and proteolytic specificities (40), (44), and (109).

Serine proteases:

Serine proteases are characterized by the presence of a catalytic triad composed of amino acids serine (nucleophile), aspartic acid (electrophile), and histidine (base) in their catalytic site. In this case serine acts as a very active residue and forms the covalent bond with the substrate (40). These proteases are found among viruses, bacteria and eukaryotes, suggesting

that they are important to the organisms. Based on their structural similarities, these proteases are grouped into 20 families, which are further subdivided into about six clans with common ancestors (108), (107). Most of the serine proteases are active at neutral and alkaline pH, with an optimum pH 7 and 9. Based on their structure serine proteases fall into two broad categories: chymotrypsin-like and subtilisin-like (73). These proteases are responsible for co-ordinating various physiological functions in humans, including digestion, immune response, blood coagulation and reproduction (44). These proteases can be inhibited by small organic compounds (organophosphates, heterocyclic derivatives) or by protein protease inhibitors (serpins) (87).

Aspartic proteases:

Aspartic acid proteases are also known as acidic endopeptidases. The active site of these proteases consists of two aspartic acid residues. These proteases are grouped into three families, pepsin (A1), retropepsin (A2), and enzymes from pararetroviruses (A3) (106) (104), and have been placed in clan AA. Most of the aspartic acid proteases show maximal activity at low pH (pH 3 to 4). The active-site aspartic acid residue is situated within the motif Asp-Xaa- Gly, in which Xaa can be Ser or Thr. Pepstatin can inhibit the aspartic proteases (30). They are also sensitive to diazoketone compounds like diazoacetyl-DL-norleucine methyl ester (DAN) and 1,2-epoxy-3-(p-nitrophenoxy) propane (EPNP) in the presence of copper ions (102).

Cysteine proteases:

Cysteine proteases occur in both prokaryotes and eukaryotes. They are also known as thiol endopeptidases and are characterized by the catalytic dyad: cysteine (Cys), and histidine (His). In this dyad cysteine is the important active residue which involves in the formation of intermediate complex with substrates. The position of Cys and His (Cys-His or His-Cys)

residues differ among the families (107). Cysteine proteases are divided into four groups on the basis of their side chain specificity: (i) pepsin-like, (ii) trypsin-like with preference for cleavage at the arginine residue, (iii) specific to glutamic acid, and (iv) others. The best known cysteine protease is papain. Most of the cysteine proteases are active at neutral pH, very few of them are maximally active at acidic pH e.g., lysosomal proteases. Epoxides, microbial inhibitors and the compounds that block thiol groups can inhibit the cysteine endopeptidases (87).

Metalloproteases:

Metalloproteases are diverse of the catalytic types of proteases characterized by the presence of a divalent metal ion for their activity (104). These proteases contain a metal ion, generally zinc (Zn^{2+}) in their active site which is an integral part of the protein structure, while a few contain Mg2+, Ni2+, or Cu2+. Catalytic metal ion activates the water molecule, which acts as a nucleophile in catalysis. The majority of metalloproteases contain one catalytic metal ion, and those containing two catalytic metal ions so far described are all exoproteases (7). In addition to a Zn^{2+} ion in the active site, there will be a glutamic acid (Glu), or often associated with tryptophan (Trp) or histidine (His) (87). Metalloproteases are produced by all species of plants, animals, and microorganisms. Some of them function in the cell or on the membrane and some of them are secreted to the periplasm or outside the cell are called extracellular metalloproteases. Extracellular metalloproteases in mammals, matrix metalloproteases are related to many diseases. Collagenases, some snake venoms and thermolysin form bacteria are some of the examples of these proteases (46), (90), (124), and (137). They are inhibited by ion-chelating agents, some are inhibited by phosphoramidon. Matrix endopeptidases are inhibited by TIMP's (tissue inhibitors of metallo-endopeptidases) and IMP's (inhibitors of metallo-endopeptidases). The optimum pH for these proteases is from 7 to 9 (87).

Classification of proteases based on the molecular structure and homology (Barrett & Rawlings classification of proteases)

In 1993, Neil D. Rawlings and Alan J. Barrett proposed another classification of proteases that is based on the relationship between molecular structure functional homology. This system mainly depends on the availability of data for amino acid sequences and the three-dimensional structures that were developed only in the early 1990s. According to this classification system, peptidases are assigned to families and the families are again grouped in clans. A peptidase database called MEROPS was launched by Rawlings and Barrett in 1999 and the classification is also known as "MEROPS classification".

The term "family" defines a group of homologous proteases. Proteases that show significant similarity in the amino acid sequence to the type protease of the family or to another protein that has been already shown homologous to the type protease are grouped into one family. At least the relationship must exist in the peptidase unit. A family can contain only one peptidase if no homologues exist, and a single gene product like viral polyprotein containing more than one protease can be assigned to a different family. Each family is given by a letter representing the catalytic type of proteases (S, C, A, M, T, G or U, for serine, cysteine, aspartic, metallo, threonine, glutamic or unknown, respectively) and a special number, e.g. S1, M1. Apart from this, some families are even divided into subfamilies because there is evidence of a very ancient divergence within the family, e.g. S1A, S1B etc.

A "clan" comprises of all the proteases that have been arisen from a single evolutionary origin. It describes one or more families that show evidence of their evolutionary connection by their similar tertiary structures, or when structures are not available, by the order of catalytic-site amino acids in the polypeptide chain and often by common sequence motifs around the catalytic amino acids. Each clan is recognized with two letters, the first

representing the catalytic type of families included in the clan (with the letter "P" being used for a clan containing families of more than one of the catalytic types serine, threonine and cysteine). Some families cannot be assigned to clans because of the requirement of formal assignment. Such a family is described as belonging to clan A-, C-, S-, M-, T-, G- or U-, according to the catalytic type. Some clans are divided into subclans on the basis of the evidence of a very ancient divergence within the clan, for example MA(E), the gluzincins, and MA(M), the metzincins (105), (7) .

The full information about families and clans for MEROPS classification can be found at the following linkages, http://merops.sanger.ac.uk/cgi-bin/family_ index?type=p , and http://merops.sanger.ac.uk/cgi-bin/clan_index?type=p. The MEROPS Ids of proteases that will be focussed on in this dissertation are: norovirus 3C-like protease (C37.001), poliovirus 3C protease (C03.001), and TGEV main/3C protease (C30.004). MEROPS classification of proteases is showed using aspartic acid and cysteine proteases as examples in **Table 1.2**

Protease terminology

Proteases are macromolecules with a number of atoms and side chains that alters shape and charges on their surfaces and/or inside the protein. A protease identifies its peptide substrate based on its backbone conformation, amino acid sequence and position of the amide bonds being cut. The protease substrate binds to a protease in a complementary manner based on size, shape and electrostatics. In 1967, Schechter and Berger proposed a nomenclature of protease and substrate that is commonly used to number the active site of a protease. This nomenclature states that the amino acid residues of protease substrates are regarded comparative to the site of backbone cleavage. The amino acid residues towards the N-terminal from the cleavage site are said to be on the P side, and those residues towards the C-

terminal are said to be on the P' side. Right away N-terminal to the cleavage site is residue P1, and moving more N-terminally are residues P2, P3, P4, and so on. Residue right away C-terminal to the cleavage site is said to be P1', and moving more C-terminally are residues P2', P3', and so on. In this way, the protease cleaves the amide bond that links amino acid residues P1 and P1'.

Proteases contain binding pockets called substrate binding pockets for substrate amino acid residues. Those binding pockets that bind to the P side of the substrate are called S pockets, whereas those pockets that bind to the P' side of the substrate are called S' binding pockets. The protease pocket that interacts directly with the P1 amino acid residue of the substrate is called S1 pocket and the remaining pockets that bind to residues P2, P3, P4, and so on are called S2, S3, S4, and so on respectively. In the same way, the protease pocket that interacts directly with the P1' amino acid residue of the substrate is called S1' pocket, and the remaining protease pockets that bind to residues P2', P3', P4' and so on are called S2', S3', S4' and so respectively. Thus, the amino acid residues involved in the proteolysis are positioned between the S1 and S1' substrate binding pockets. The character of these pockets decides the degree of specificity of the protease (120), (141). The substrate and substrate binding pockets are shown in **figure 1.2**

Proteases and proteolytic reactions or mechanisms

Proteases breakdown peptide bonds between amino acids in proteins and peptides by the process of hydrolysis. Even though, they catalyze the same kind of reactions, a large number of proteases exist to hydrolyze the specific substrates in the precise biological environment. According to MEROPS peptidase database, there are more than 2000 different proteases exist. Catalytic mechanisms of four groups of proteases (serine, cysteine, aspartic,

and metallo) that are intensively studied for their catalytic mechanism and substrate specificity are described below.

Serine protease catalytic mechanism

Serine proteases are enzymes that cleave peptide bonds in proteins, in which serine acts as the nucleophilic amino acid at the active site. The active site of serine protease consists of three amino acids serine, histidine, and aspartic acid forming a catalytic triad. The basic mechanism of action of serine proteases involves transfer of the acyl portion of a substrate to a functional group of the enzyme (44). The two basic steps of catalysis by this group of enzymes thus include:

(i) Firstly, the formation of an ester bond between the oxygen atom of serine and the acyl portion of the substrate - which produces a tetrahedral intermediate and releases the amino part of the substrate; and

(ii) Secondly, the attacks of water molecule on the acyl-enzyme intermediate, which breaks it down and releases the acidic product - while regenerating the original enzyme form **(Figure 1.3)**.

Cysteine protease catalytic mechanism

Cysteine proteases are enzymes that cleave peptide bonds in proteins, in which cysteine acts as the nucleophilic amino acid in the active site of the protease. The active site of cysteine protease consists of two amino acids cysteine, and histidine forming a catalytic dyad. But recently, there are some studies reporting that there is a third amino acid that would be potentially glutamic acid in the active site of cysteine proteases (49). The basic mechanism of cysteine proteases involve:

(i) The first step is deprotonation of cysteine in the enzyme's active site by an adjacent amino acid with a basic side chain, usually a histidine residue. Then, the correct positioning of the substrate within the catalytic site cleft brings the scissile bond into sufficiently close proximity with the cysteine to facilitate nucleophilic attack by the thiolate on the carbonyl C of the scissile peptide to generate an unstable, covalently bound tetrahedral intermediate.

(ii) During nucleophilic attack on the substrate carbonyl group, the shift of negative charge onto the carbonyl O favours the formation of oxyanion hole, which helps in stabilizing the tetrahedral intermediate formed.

(iii) Collapse of the tetrahedral intermediate complex is promoted by histidine by protonating the substrate amide N. This leads to the release of C-terminal portion of the substrate by forming an acyl-enzyme intermediate and the amide product.

(iv) The N-terminal portion of the substrate is released by the hydrolysis of the acyl-enzyme, which is achieved by the diffusion of the water molecule into the catalytic site that is followed by nucleophilic attack on the substrate carbonyl carbon, promoted by histidine and potentially glutamic acid (87), (107) **(Figure 1.4)**.

Aspartic acid protease catalytic mechanism

Aspartic proteases are a family of proteases that use an aspartic acid residue for catalysis of their peptide substrates. Their catalytic site consists of two aspartic acid residues. The enzyme-substrate intermediate formation is of the general acid-base type rather than the covalent type. The basic mechanism is illustrated in the figure (the model chosen is pepsin):

(i) The catalytic groups of the two aspartic acid molecules at the active site (aspartic acid 32 and 215) are linked through a hydrogen bond. In case of aspartic acid

proteases, nucleophilic attack is facilitated by two simultaneous proton transfers: the first proton is transferred from a water molecule to the carboxylate ion of aspartic acid 32, and the second proton from the carboxyl group of aspartic acid 215 to the oxygen atom of the carbonyl group of the substrate. These transfers of protons lead to the formation of a neutral tetrahedral intermediate.

(ii) The intermediate is destroyed by the similar dual transfer mechanism, which donated a proton to the carboxylate ion of aspartic acid 32, thus hydrolysing the C-N bond of the substrate by transferring the proton from aspartic acid 215 carboxyl group to the nitrogen atom (87), (106) **(Figure 1.5)**.

Metalloproteases catalytic mechanism

Metalloproteases contain a metal ion in their active site, generally zinc (Zn^{2+}), which is an integral part of the protease. These proteases do not form covalent enzyme-substrate complex intermediates. Apart from Zn^{2+} ion, the active site always contains a glutamic acid residue (Glu), sometimes associated with tryptophan (Trp) or histidine (His). The catalytic mechanism of metalloproteases is shown in the **(Figure 1.6)**. The basic catalytic mechanism of metalloproteases involves:

(i) The nucleophilic attack on the carbonyl carbon by zinc bound to water is facilitated by the carboxylate ion of the glutamic acid at the active site.

(ii) The transfer of proton from the carboxylate ion of the glutamic acid to the eliminated NH_2R' group breaks down the tetrahedral intermediate, finally releasing carboxylic acid with the aid of water molecule (87).

The proteases that will be focussed on in this dissertation are cysteine proteases because all the three viral proteases such as norovirus protease, poliovirus protease, and

transmissible gastroenteritis virus protease contain cysteine as active nucleophile in the catalytic site.

Proteases as targets in drug discovery

Proteases are very crucial enzymes that account for ~2% of the genes in humans, infectious organisms like bacteria, viruses, and fungi, and other forms of life. They control most physiological mechanisms by regulating the activation, synthesis and turnover of all proteins. Thus, they are the central regulators of conception, birth, growth, maturation, ageing, diseases and death. Proteases are also important for replication or transmission of infectious disease causing agents such as viruses, bacteria, and parasites (1). In case of viruses, the proteases are required for the processing of precursor polyproteins to yield functional viral proteins and therefore viral proteases presents themselves as potential viable targets for the development of antiviral strategies (132). On the basis of their importance in health and disease, protease inhibitors have already been developed into blockbuster drugs and diagnostics, and many other protease inhibitors are in clinical trials. Many protease inhibitors have showed assured therapeutic activities in preclinical trials in animals and in early clinical trials in man for viral, bacterial, and parasitic infections (1). For example, many inhibitors are developed against HIV proteases and this approach has been one of the primary means for treating HIV infections (52), (111). A review on "protease inhibitors in the clinic" has summarized many more protease inhibitors in clinical development.

The proteases that are focussed in this dissertation are viral proteases from norovirus, poliovirus, and porcine transmissible gastroenteritis virus. These three viruses are highly contagious and there is a need to develop antiviral compounds against these proteases to control the outbreaks and transmissions. Although vaccines are available for poliovirus, there is a need for development of antiviral drugs, which is discussed under poliovirus literature

review. The additional advantage in developing viral protease inhibitors is that the inhibitor can act against proteases of different strains in the same family of virus, thereby not only controlling the disease during an outbreak, but also reducing the transmission (146).

CHAPTER 2-VIRAL PROTEASES AND DISEASES

LITERATURE REVIEW

Viruses are small infectious agents responsible for causing a variety of serious diseases in humans, other animals, and plants. As viruses are clinically and scientifically important, they have been under intensive study from a long period of time ever since their first isolation about a century ago. Viral infection has become a real global threat. Viral infection is not only related to serious human diseases like Swine flu, AIDS, SARS, ebola and avian influenza (27), but is also associated with common human diseases including common cold, liver carcinoma, chickenpox, shingles, cold sores, gastroenteritis and etc. Viral proteases are the main enzymes that are involved in some virus maturation and thus represent an attractive target for the development of novel antiviral agents (132). Some viral proteases and related diseases are shown in the below Table.

Viral proteases and their related diseases:

Protease	Related diseases
HIV protease	AIDS
Norovirus protease	Acute gastroenteritis
Poliovirus 3C protease	Poliomyelitis
TGEV main protease	Severe and often fatal diarrhoea in young pigs
Hepatitis C Virus protease	Liver injury, liver cancer and cirrhosis
Hepatitis A Virus 3C protease	Acute liver infection

Proteases	Related Disease
Epstein-Barr virus protease	Lymph proliferative infection
Rhinovirus 3C protease	Common cold
Cytomegalovirus protease	Retinitis
Dengue Virus NS2B-NS3 protease	Dengue fever
Varicella Zoster virus protease	Chickenpox, shingles
Herpes simplex virus protease	Oral lesions, genital herpes
West Nile Virus NS2B-NS3 protease	Encephalitis
FMDV 3C protease	Foot and mouth disease in animals

Norwalk virus

Noroviruses are the most common cause of epidemic non-bacterial gastroenteritis, responsible for approximately 90% outbreaks worldwide, and a major cause of foodborne illness (72). In the United States alone approximately 21 million illnesses attributable to norovirus are estimated to occur annually and they are responsible for more than 50% of all foodborne outbreaks of gastroenteritis (139), (154). The virus is transmitted by the fecal-oral route through contaminated food and water, by direct contact, and possibly by transmission through air-borne viral particles. Noroviruses are highly infectious and just a few particles can cause infection (21). Outbreaks have been reported in a variety of settings including hospitals, military troops, cruise ships, schools, prisons and childcare centres (60). When a person becomes infected with norovirus, the virus begins to multiply within the small

intestine. After approximately 1 to 2 days norovirus symptoms can appear. The principal symptom is acute gastroenteritis that develops between 24 and 48 hours after exposure, and lasts for 24 to 60 hours. The disease is usually self-limiting, and characterized by nausea, forceful vomiting, watery diarrhea, and abdominal pain; and in some cases, loss of taste, general lethargy, weakness, muscle pains, headache, coughs, and low-grade fever may occur. The median incubation period is classically 32 hours, with a range of 24-48 hours, but can extend outside this range (12-50 hours). Symptoms usually resolve within 24-48 hours but may range from 12-72 hours. However, in some people symptoms can last longer than previously thought, particularly among the elderly and in young children, as well as among transplant recipients and the immunosuppressed, such as people on immunosuppressive therapy, or ill with immune-modulating diseases (e.g. HIV) (155). It appears that individuals develop short-term immunity following infection and immunity is strain specific. The genetic variability in circulating noroviruses indicates that individuals are likely to be repeatedly infected during their lifetime. Short-lived immunity may explain in part the high attack rates in all age groups in an outbreak (77). There is evidence suggesting to norovirus infection to be based on histo-blood group antigens (131). Viral shedding is greatest during the acute illness and the amount of virus excreted decreases rapidly with recovery. Viral shedding in stools is greatest over the first 24-48 hours (28).

Noroviruses are a group of single-stranded RNA viruses that have been classified as members of the family Calciviridae (97), (118). Comparisons of genome sequence and organization have placed these viruses in the family Caliciviridae: a viral family consisting of five currently recognized genera, one of which is noroviruses (91). Noroviruses have been subdivided into five genogroups (GI-GV) with the noroviruses infecting humans most commonly occuring in genogroups I and II. The GI and GII noroviruses are further subdivided into genotypes. For example genogroup II, the most prevalent human genogroup,

presently contains 19 genotypes. Genogroups I, II and IV infect humans, whereas genogroup III infects bovine species and genogroup V has been recently isolated in mice (100). Noroviruses commonly isolated in cases of acute gastroenteritis belong to two genogroups: genogroup I (GI) includes Norwalk virus, Desert Shield virus, and Southampton virus and genogroup II (GII), which includes Bristol virus, Lordsdale virus, Toronto virus, Mexico virus, Hawaii virus and Snow Mountain virus (152).

The Norwalk virus (NV) is responsible for an outbreak of non-bacterial gastroenteritis in a school in Norwalk, Ohio, in 1968 and it is the first enteric calicivirus discovered in 1972 using immune electron microscopy (54). The genome of the Norwalk virus is the first to be sequenced using hybridization assays (143). Among all the genogroups of noroviruses, Norwalk virus is the most studied prototype virus and is classified as GI/1 strain. In recent years, 75-100% of norovirus outbreaks and sporadic cases worldwide are due to the strains from genogroup II (92). Globally 60-70% of all norovirus outbreaks are associated with GII/4 strains (125).

Currently, there is no vaccine or antiviral therapy available for the control of norovirus infections due to the lack of permissive cell culture systems and animal models for human noroviruses. Noroviruses show high genetic diversity, and because of that immunity to one strain cannot provide protection from another strain (35). According to National Institute of Allergy and Infectious Diseases (NIAID) ,noroviruses are classified as category B pathogens due to their highly contagious nature and a potential to cause serious public health issue, because as few as 10-100 virions are enough to start infection in a healthy individual (60). Hence there is a need for the development of antiviral drugs to control the norovirus infections.

Norovirus genome consists of a molecule of single-stranded positive sense RNA of approximately 7700 nucleotides with a polyadenylated 3' terminus (68), (69). These are non-enveloped viruses. The genome of norovirus consists of three open reading frames (ORFs); ORF1 is located towards the 5' end and encodes a large 200 kDa non-structural precursor polyprotein, ORF2 encodes the major capsid protein (VP1), and ORF3 encodes a minor structural protein (VP2) which helps in the assembly of newly synthesized viruses (10) **(Figure 2.1)**.

In vitro translation and mutagenesis studies indicated that the 200kDa ORF1 polyprotein of noroviruses (Norwalk virus) is cleaved by the action of a 3C-Like protease to initially generate three separate functional protein products. Full processing of the ORF1 precursor polyprotein produces six mature non-structural protein products (11), (60) : an N-terminal protein (p45), an NTPase (p40), a 3A-like protein (p22), a Vpg protein (p16), the 3C protease (p19), and an RNA-dependent RNA polymerase (p57) **(Figure 2.2)**. The protease also allows the preferential expression of viral proteins by inhibiting the translation of host proteins by cleavage of the poly (A) binding protein (66). The scissile bonds cleaved by the norovirus protease within its 200 kDa polyprotein substrate are within the dipeptide recognition sequences Glu/Gly or Gln/Gly, where the "/" indicates the cleavage site. The protease preferentially cleaves at LQ/GP (p45/p40), LQ/GP (p40/p22), PE/GK (p22/p16), FE/AP (p16/p19), and LE/GG (p19/p57) junctions to generate mature proteins (11) (60). The protease preferentially accommodates a glutamine residue at the active site S1 position and a glutamate residue will also be accepted (38). The catalytic site of Norwalk virus protease consists of Cys139, His30 and Glu54 as active amino acid residues. Cys139 acts as nucleophile in the active site **(Figure 5.1)**. Norwalk virus protease shows similar chymotrypsin-like fold as picornaviral proteases.

Comparative studies of norovirus protease amino acid sequences with those of poliovirus (PV), human rhinovirus (HRV), hepatitis A virus (HAV), Chiba virus (CV), Southampton virus (SV) and foot and mouth disease virus (FMDV) showed that the norovirus protease belongs to the family of chymotrypsin like cysteine proteases because of its similarity to picornaviral 3C protease. This is confirmed by the availability of two X-ray crystal structures of the protease from group 1 norovirus (85). The active site of cysteine proteases generally possess an active site dyad consists of cysteine and histidine, but the three dimensional structures of the PV, HRV, HAV, FMDV, CV and SV 3C proteases showed a third amino acid in the active site, which may be either aspartate or glutamate (85), (48). Since norovirus protease is essential for the processing of the 200 kDa precursor polyprotein to produce mature or functional viral proteins, the viral protease presents itself as an attractive target for the development of antiviral strategies.

Poliovirus (PV)

Poliovirus is a member of the genus Enterovirus, which belongs to the family Picornaviridae. The family Picornaviridae is a large family of small, non-enveloped, positive stranded RNA viruses which include a number of important human and animal pathogens such as poliovirus (PV), hepatitis A virus (HAV), Coxsackievirus (CV), human rhinovirus (HRV) and foot and mouth disease (FMDV). The family of Picornaviridae consists of nine genera namely Enterovirus, Rhinovirus, Cardiovirus, Aphthovirus, Hepatovirus, Parechovirus, Erbovirus, Kobuvirus, and Teschovirus. Polioviruses belong to the genus Enterovirus and the genus can be further subdivided into eight clusters consisting of Poliovirus, Human enterovirus A, Human enterovirus B, Human enterovirus C, Human enterovirus D, Simian enterovirus A, Bovine enterovirus, and porcine enterovirus B (23). Poliovirus, Coxsackieviruses, and echoviruses are the major pathological disease causing viruses in the genus Enterovirus. These viruses are responsible for causing diseases such as meningitis, encephalitis,

myocarditis, paralysis, diarrheic and respiratory diseases (84). Enteroviruses replicate mainly in the gastrointestinal tract and an estimated infection per year worldwide is one billion (88).

Initially human enteroviruses were classified into four categories based on the clinical manifestations observed in human infections and experimentally inoculated mice. The four categories are (i) polioviruses, responsible for causing flaccid paralysis in humans but not in suckling mice without CD155; (ii) coxsackie A viruses (CAV), related to human central nervous system pathology and skeletal muscle inflammation and flaccid paralysis in suckling mice; (iii) coxsackie B viruses (CBV), connected with illness of the human cardiac and central nervous systems, focal lesions in skeletal muscle, brain, and spinal cord, and as well as spastic paralysis in suckling mice; and (iv) echoviruses, not connected with human disease nor with paralysis in mice (23). Highly sophisticated techniques available in the field of molecular biology lead to the proposal of a new modified classification under which human enteroviruses are subdivided into five species consisting of poliovirus, and Human enterovirus A, B, C, and D. As polioviruses show greatest similarity of relation to the C-cluster human enteroviruses (CAV11, CAV17, and CAV 20) on the basis of genome sequences, recently the Executive Committee of the International Committee on Taxonomy of Viruses (ICTV) recommended a proposal to move the polioviruses into the Human enterovirus C species. In fact polioviruses differ from these three C-cluster coxsackie A viruses in the structural capsid region (P1) (16).

Poliovirus is responsible for causing poliomyelitis in humans. Poliomyelitis or often called polio or infantile paralysis is an acute central nervous system disease. The epidemics of the poliomyelitis began to occur in the United States and Europe during the beginning of the 20th century (99). In 1840, Jakob Heine first recognised the poliomyelitis as a distinct condition (93) and in 1908 poliovirus was identified as the causative agent for poliomyelitis by Landsteiner and Popper (99). Even though the epidemics of poliomyelitis began to occur

near the beginning of the 20th century, there is an historical record stating that isolated cases of poliomyelitis have been occuring for many years (86). Poliovirus shows structural similarity with other human enteroviruses (Coxsackieviruses and echoviruses), as well as to human rhinoviruses. In common they all use immunoglobulin like molecules to recognise and enter into host cells (42). A study of phylogenetic analysis of the RNA and protein sequences of poliovirus suggests that, a mutation within the capsid may have helped in the evolution of poliovirus from a C-cluster Coxsakie A virus ancestor (51). Based on slight differences in the capsid protein, poliovirus is divided into three serotypes, Poliovirus 1 (PV1), poliovirus 2 (PV2) and poliovirus 3 (PV3). The cellular receptor specificity and antigenicity are determined by the capsid proteins. All three forms of poliovirus are highly infectious, but among them PV1 is the most infectious form in the nature (93). Wild poliovirus type 1 is highly restricted to the regions of India, Pakistan, Afghanistan, and Egypt, but it is highly prevalent in West and Central Africa. Wild poliovirus type 2 was last identified in October 1999 in Uttar Pradesh, India and probably it has been extirpated. Wild poliovirus type 3 is confined to five countries during 2004 endemic namely Nigeria, Niger, Pakistan, India, and Sudan (58).

Poliovirus is an enterovirus and the virus is transmitted person to person by fecal-oral route and also oral-oral route. Poliomyelitis is highly contagious and wild polioviruses in endemic areas can infect the entire population. Poor living conditions, tropical and subtropical conditions, high birth cohorts, and high population compactness are the important hazardous circumstances involved in the transmission of poliovirus. When a person becomes infected with poliovirus, the replication of virus occurs in the alimentary tract (84), (86), and (58). The symptoms are sometimes inconspicuous or only mild. The virus is shed in the feces of the infected individual. For any virus, the primary determinant of infections is its capacity to enter a cell and generate additional infectious particles. The presence of CD155 receptor

on the cells and tissues of humans and a few subhuman primates (chimpanzees and Old World monkeys) is an opinion to state that, the virus can infect those cells and tissues. But however, poliovirus is strictly a human pathogen and does not naturally infect any other species, even though subhuman primate species can be experimentally infected (99), (84), and (83). In most of the cases of poliovirus infection, 95% of cases show only a primary, temporary presence of viremia and the poliovirus infection may be asymptomatic. In some cases, about 5% the virus enters and multiplies in other places such as brown fat, reticuloendothelial tissue and muscle. The prolonged viral multiplication gives raise a condition called secondary viremia, which leads to the advancement of onset symptoms such as fever, headache and sore throat (117). Less than 1% of poliovirus infections lead to paralytic poliomyelitis, the condition in which the virus enters the central nervous system and replicates in motor neurons within the spinal cord, brain stem. The symptoms of paralytic poliomyelitis include an asymptomatic phase of 1-3 days, followed by acute onset of flaccid paralysis with fever. Depending upon the site of viral replication in the central nervous system, the paralytic poliomyelitis may affect skeletal muscles causing spinal poliomyelitis, respiratory muscles causing bulbar poliomyelitis, or may affect both muscles causing bulbo-spinal poliomyelitis. The loss of motor neurons leads to temporary or permanent paralysis (58).

Poliomyelitis is a seasonal disease with high transmission rates during summer and autumn and low transmission rate in winter. Paralytic assail is based on the serotype and wild poliovirus type 1 is connected with highest rates of paralytic attack (~0.5%) and lowest rates of paralytic attack are connected with poliovirus type 2 (<0.05%). The paralytic attack of poliovirus is thought to be rare and an accidental distraction of gastrointestinal infection (84). The process by which the virus reaches the central nervous system is insufficiently understood. Three non-mutually exclusive hypotheses have been propounded to explain the

process of entry of poliovirus into the central nervous system. Primary viremia is common in all the three theories. The first theory states that the virus may enter the central nervous system from the blood by crossing over the blood brain barrier, without the influence of its receptor CD155 (149). The second hypothesis, explains that the virus is transported to the central nervous system ascending from the muscle to the spinal cord and brain through nerve pathways by retrograde axonal transport (89), (112). The third hypothesis, explains that there may be a chance of spreading the virus to the central nervous system by infected primary monocytes or macrophages expressing CD155 (23).

The genome of poliovirus is approximately 7400 nucleotides in length and composed of single stranded positive sense RNA consisting three physically separate regions. A relatively long 5' non translated region (5'NTR) of 742 nucleotides , a single open reading frame (ORF) which codes for a polyprotein of 2209 amino acids and a considerably short 3' non translated region (3' NTR) of 70 nucleotides followed by a virus-encoded poly(A) tract of 60 adenine residues length (25), (62), (98), and (128). The 5' NTR region of the poliovirus genome is subdivided into (i) 5' terminal clover leaf, which acts as an indispensable cis-acting factor in viral replication and in regulating the process of translation initiation; and (ii) the internal ribosome entry site (IRES), for its propensity to initiate the process of cap-independent translation of the viral mRNA regardless of its location within an RNA molecule and even a free 5' end (84), (23). Comparatively there is less known about 3' NTR in contrast to 5' NTR. The 3' NTR region is polyadenylated and there is evidence that it has an important role in RNA replication. The deletion of the 3'NTR showed only minimum effects on the propagation of poliovirus in HeLa cells. But in the cells of neuronal origin, the ability of the virus propagation is evidently reduced both in vitro and in vivo (16), (15). The 250 kDa polyprotein is encoded by the single open reading frame (ORF). The single open reading frame (ORF) can be subdivided into regions P1, P2, and P3, which encodes the structural and

non-structural proteins. The unstable polyprotein is cleaved by the proteolytic activity of virus-encoded proteases, 2Apro and 3C/3CDpro in cis and trans, finally producing the functional proteins required for viral proliferation (39) **(Figure 2.3)**.

Poliovirus life cycle occurs within the limits of the cytoplasm of infected cells. Attachment of the poliovirion to its cell surface receptor, the human poliovirus receptor (hPVR) or CD155 initiates the life cycle of poliovirus (63), (78). Binding of poliovirus with CD155 provides an irreversible conformational change in the viral particle required for its entry into the cell (55), (145). The definite mechanism used by the poliovirus to release its genomic RNA into the infected host cells is not well established. Once the virus is attached to its cell surface receptor CD155 on host cell membrane, the process of release of viral genomic RNA into the host cell is thought to occur by one of the two pathways: (i) RNA is injected into the host cell cytoplasm through the formation of a pore in the plasma membrane; or (ii) the virus is taken up by receptor-mediated endocytosis, because the virus can be found on endosomes (23). The latter hypothesis is supported by experimental evidence and it suggests that the virus particle binds with CD155 and is taken up via endocytosis. Once the virus is internalized, the genomic RNA is released (14).

When a cell is infected with poliovirus, the genomic RNA is released into the cytoplasm of the host cell. This virion RNA functions as messenger RNA and is translated into a single autocatalytic polyprotein. The proteolytic cleavage products of this polyprotein serve as capsid precursors and replication proteins (140). Poliovirus shows a certain degree of posttranslational control of the amounts of viral proteins generated via an intricate cascade of fast and slow proteolytic cleavages by its three proteases 2Apro, 3C/3Dpro (84). Both 2Apro and 3Cpro are autocatalytic in nature and they can release themselves from the polyprotein by self-cleavage (136). Once the process of translation is completed, the processing of polyprotein takes place in three stages. First, P1 precursor is released from the nascent P2-P3

protein by 2Apro in a cis-cleavage manner. Second, 3CDpro releases the P3 from the remaining protein through cis-cleavage (70). The non-structural proteins 2A, 2BC, 3AB, 2B, 2C, 3A, 3B (VPg), 3Cpro, 3Dpol and the capsid proteins VP0, VP1, and VP3 are released from their precursors by the trans-cleavage action of the 3CDpro. The 3Cpro encoded by poliovirus carries out a main cleavage between 2C and 3A (fig). Specifically 3Cpro/3CDpro cleaves at a Glutamine-Glycine pair, whereas, 2Apro cleaves at a Tyrosine-Glycine pair (84). Other than polyprotein processing, 2Apro is involved in shutting off the cap-dependent host cell translation by cleaving the complex eIF-4F and 3Cpro inactivates transcription factor TFIIIC and cleaves the TATA box binding protein, there by inhibiting the cellular transcription (20), (64).

The poliovirus 3Cpro is the most specific of its class, cleaving solely at glutamine↓glycine amino acid pairs ("↓" indicates the cleavage site). The refined X-ray crystallographic structure of poliovirus 3C protease was determined by Mosimann et al in 1997. Comparative crystal structural analysis of 3C proteases from hepatitis A virus (HAV), human rhinovirus (HRV), and poliovirus (PV) showed that these proteases show a two-domain β-barrel fold similar to that of chymotrypsin supporting the fact that these viral proteases are basically different from that of classical cysteine proteases, like papain (119). The 3C protease of poliovirus consists of two six-stranded antiparallel β-barrel domains and it shows structural similarity to the chymotrypsin-like serine proteases. The translated polyprotein is proteolytically cleaved at 9 of 12 specific processing sites by poliovirus 3C protease and is mainly responsible for the production of separate gene products. The shallow active site cleft of the poliovirus 3Cpro is larger than the active sites of the chymotrypsin-like serine proteases and it is located at the junction of the two β-barrel domains. The active site of poliovirus 3C protease consists of a catalytic triad formed by Cys147 which acts as

nucleophile, His40 which acts as general acid/base catalyst and Glu71 **(Figure 5.2)**. The oxyanion hole consists of the residues Gly145, Gln146, and Cys147 (65), (82).

Although poliovirus infection is no longer a major public threat in the developed world, poliovirus continues to be one of the extremely studied and most excellent understood model viruses to date (84). The poliovirus field is lucky to have more than one excellent vaccine, because both of these vaccines showed their potency in the production of anti-poliovirus antibodies which provides protection from the disease (74). The first polio vaccine was developed in 1955, namely Salk polio vaccine or inactivated polio vaccine (IPV) by Jonas Salk on April 12. IPV was shown to effectively immunize and protect against poliomyelitis. Then, in 1963 the live attenuated oral polio vaccine (OPV) or Sabin polio vaccine was developed by Albert Sabin which was demonstrated to be both safe and effective (23). Both of these vaccines have their own advantages and disadvantages. In 1988, a partnership led by the World Health Organization (WHO), Rotary International, the US Centers for Disease Control and Prevention (CDC) and the United Nations International Children's Emergency Fund (UNICEF) launched a project to eradicate poliovirus from the world namely The Global Polio Eradication Initiative (GPEI). By the end of 2007, the numbers of poliomyelitis cases were almost reduced and only four countries Afghanistan, India, Nigeria, and Pakistan, remain with endemic poliovirus transmission (22).

The Global Polio Eradication Initiative (GPEI) is highly depended on oral polio vaccine (OPV), which is economical and easily administered live, attenuated vaccine. Even though OPV is normally safe and effective, in some parts of India, especially in Uttar Pradesh and Bihar, it failed to provide protection against poliomyelitis. OPV itself can cause vaccine-associated paralytic polio (VAPP) about 1 in 750,000 vaccines (130). Generally, healthy vaccinated individuals can excrete the OPV viruses in the stool for up to 8 weeks and these vaccine derived polioviruses (VDPVs) circulating among non-immune persons in poorly

vaccinated communities have the capacity to produce neurovirulence and cause poliomyelitis (76). IPV is safer than OPV and it does not carry any risk of generating VDPVs or causing poliomyelitis. But it is much more expensive than OPV and this will create a problem to its universal use by the world's poorest countries. IPV must be administered parenterally and to induce humoral immunity, it requires a minimum of two doses that are spaced ideally at least 2 months apart, and it will not provide sufficient mucosal immunity to protect against poliovirus replication and virus shedding and transmission. IPV is far from ideal, because it may require altered dose regimens, adjuvants to improve its effectiveness and alone it is not sufficient to control an outbreak (22). Hence there is a clear need to develop new weapons with increased ability and effectiveness to control a polio outbreak in the post-eradication era. Development of antiviral drugs is an excellent solution to control the poliovirus outbreaks. Recently, National Research Council (NRC) conducted a workshop on considering the role of antiviral drugs in a polio outbreak entitled "Exploring the Role of Antiviral Drugs in the Eradication of Polio". According to the report presented by NRC, the estimated cost of first antiviral drug development will be $75 million which is very less when compared to the money spent on the GPEI (approximately $5.3 billion) and it is estimated that about $2 billion more will be needed to finish the project. Development of antiviral agents for polio depicts an excellent cost-effective constituent to assure a successful reverts on investment (22).

The concept of antiviral therapy is not new for poliovirus infection. This concept of antiviral treatment for poliovirus infection became faded by the development of polio vaccines. Later in the 1950s, after the achievement of poliovirus propagation in cell culture, many compounds supressing the in vitro poliovirus replication were identified (17). Even though the idea of antiviral therapy is not new for polio field, but the use of antivirals for polio is new, which has successfully achieved vaccination over the past 50 years. One such

antiviral target for poliovirus is 3C protease. The 3C proteases of picornaviruses have been recognised as targets for the structure-based drug design, because they are responsible for the majority of cleavage events (65). Recently one class of inhibitor was developed with a Michael-acceptor group at its C-terminus which bounds irreversibly to the catalytic cysteine leading to the inhibition of human rhinovirus 3C protease activity not only in vitro but also in vivo (48). Since 3C protease of poliovirus is essential for the processing of precursor polyprotein to produce mature or functional viral proteins and also to inhibit the cellular transcription, the 3C protease of poliovirus presents itself as an attractive target for the development of antiviral drugs.

Transmissible gastroenteritis virus (TGEV)

Transmissible gastroenteritis virus (TGEV) is the main cause of gastroenteritis in pigs and piglets. Transmissible gastroenteritis (TGE) is considered as highly infectious disease in pigs, because of its high mortality rate in young piglets (33). TGEV is an enterophathogenic coronavirus (CoV) of the family Coronaviridae with in the order Nidovirales. The order Nidovirales is comprised of two families, namely (i) Coronaviridae, and (ii) Arteriviridae, which all commonly synthesize a nested set of multiple subgenomic mRNAs. The members of family Arteriviridae are known only to infect animals. The family Coronaviridae is comprised of two genus (i) Torovirus, and (ii) Coronavirus of similar genome organization and replication strategies with different virus morphology and genome length. Coronaviruses are responsible for common cold in humans after rhinoviruses and they are known to cause economically important diseases in cattle, poultry, pigs and serious disease in cats (80). In humans, apart from common cold, coronaviruses cause a notable portion of upper and lower respiratory tract infections, which includes bronchiolitis, and pneumonia. Most of the animal coronaviruses are associated with enteric and respiratory diseases in livestock and domestic animals, resulting in considerable economic losses worldwide (147). The first coronavirus

isolated was infectious bronchitis virus (IBV) in 1930s, which is followed by the isolation of mouse hepatitis virus (MHV), and transmissible gastroenteritis virus (TGEV) in 1940s (33), (8).

Serologically coronaviruses are classified into three distinct groups. To date, serogroups I and II are known to infect mammals and serogroup III is known to infect birds (36). Serogroup I contains the prototypes HCoV-229E, HCoV-NL63, and many animal coronaviruses. Serogroup II contains prototypes like MHV, SARS-CoV, and multiple animal viruses. Naturally, many coronaviruses are known to infect only one animal species or very few related species. SARS-CoV is exempted from this rule, because it is known to infect a broad range of mammals, including humans, nonhuman primates, Himalayan palm civets, raccoon dogs, cats, dogs, and rodents (37), (113).

Coronavirus infection induces a variety of cytopathic effects (CPEs), depending on the virus strains and host cells. When a pig or piglet is infected with TGEV, the virus replicates in the cytoplasm of absorptive epithelial cells located in the villi of the small intestine and to a lesser extent in the respiratory tract (94), (59). TGEV transmission occurs orally and the virus is shed in the feces and in respiratory discharges of infected animals. TGEV appears in two clinical forms namely epizootic and enzootic. The replication of TGEV in its epizootic form results in the degeneration or shortening of the small intestinal villi. This reduces the absorptive capability of small intestinal villi, which results in severe watery diarrhea causing piglets to die of dehydration during the first two or three weeks of age. TGEV diarrhea in its enzootic form generally occurs in piglets older than two weeks of age and can persist through two weeks post weaning. This persistent infection of a herd with TGEV is the result of introduction of susceptible animals into the immune or partially immune herd (29), (31). TGEV is considered as prototypical enterophathogenic virus and often, it is used as a model in the studies of pathophysiology and pathogenesis of viral diarrhea (94).

Coronaviruses are also susceptible to mutations at high frequency like other RNA viruses, because of the high error frequencies of the RNA polymerases which leads to the origin of multiple mutant virus strains. Deletion mutants also arise frequently and the emergence of porcine respiratory coronavirus (PRCoV) from TGEV is the best example for deletion mutant (59). PRCoV emerged as a new virus during early 1980s in the Europe that caused widespread epizoonotic respiratory infections in pigs (103). PRCoV is a spike (S) gene deletion mutant of TGEV and the comparison of two strains suggest that tropism and pathogenicity are influenced by S gene product (spike protein) and ORF3 (59). The replication of PRCoV has been found to occur in nasal, tracheal, bronchial, bronchiolar and alveolar epithelial cells and in alveolar macrophages. PRCoV has also been detected in the United States and was first isolated in 1989 by Dr. Howard Hill (5). European and US PRCoV isolates differ slightly in their S gene and ORF3a deletions. Clinically PRCoV is non-pathogenic and field experiences have shown that herds exposed for the first time have no symptoms of disease, but however it affects lung tissue in the presence of other respiratory pathogens (156). TGEV infection starts with the binding of virus to the cellular receptor porcine aminopeptidase N (pAPN) through its spike (S) protein. In addition to its pAPN binding site, spike protein of TGEV has sialic acid binding activity located on the N-terminal region of the protein. Virus mutants lacking this sialic acid binding activity did not infect piglets showing that binding to sialoglycoproteins or mucins in the gastrointestinal tract is important for TGEV to reach the pAPN (24), (122), (9), and (123). As PRCoV is the spike gene deletion mutant of TGEV, it cannot infect the intestines of piglets. There is a study showing that PRCoV induces partial cross-protective immunity against the TGEV transmission in herds (59). Apart from mutations, another common feature of coronavirus genetics is high frequency of RNA recombination in all groups of coronavirus, including murine hepatitis virus (MHV), TGEV, and IBV. The usage of vaccines against IBV was

limited, because of the emergence of IBV variant strains from the recombination of vaccine and wild-type virulent strains (147).

Currently, there are no vaccines or antiviral therapy available for the control of human coronaviruses, but vaccines against common animal coronaviruses, such as IBV and CCoV, are generally used to prevent serious diseases in juvenile animals, and efforts are being made to improve the efficacy of these vaccines. In case of SARS-CoV, a variety of vaccines were developed using inactivated whole virus, expressed proteins, and live virus vectors and they all shown to induce the production of neutralizing antibodies in vitro. In animal models, they all have shown to decrease the viral replication and inflammation. In case of TGEV, priming of piglets with PRCoV acts as a naturally attenuated vaccine, because they both share 90% of homology in their genome sequence other than the absence of spike protein in PRCoV. This absence changes the infectivity of PRCoV from TGEV and PRCoV is no longer enterophathogenic (146), (80). There are several problems associated with coronaviruses, even though vaccines are developed against IBV, CCoV, and TGEV. Hence there is a need for the development of antiviral drugs to control the coronavirus infections. At molecular level, coronavirus infection is well understood by the availability of highly sophisticated tools and techniques in the field of molecular biology. It suggests that a number of factors like RNA-dependent RNA polymerase, virus encoded proteases, cellular receptors for binding and infection, and viral spike glycoprotein can be used as targets for the development of antiviral agents (147).

By taking the above discussed issues into consideration, it is better to develop broad spectrum antiviral drugs against the all existent pathogenic coronaviruses. However, it is not easy to develop broad spectrum inhibitors, because the development of broad spectrum antiviral agent is completely dependent on the availability of a conserved target within the entire coronavirus genus. Some of the key factors like structural proteins (S-spike, E-

envelope, M-membrane, HE-hemagglutinin esterase, and N-nucleocapsid) show considerable diversity among the coronaviruses and they are not suitable for the development of broad spectrum antivirals (142), (134), (75), and (115). Apart from structural proteins, the RNA-dependent RNA polymerase, RNA helicase, and Main protease (Mpro) are the excellent non-structural protein targets for the development of antivirals. There is no structural evidence available for the RNA-dependent RNA polymerase and RNA helicase proteins, enhancing the obstacle for designing a versatile antiviral against all coronaviruses (147). Main protease (Mpro) or 3C-like protease of coronaviruses can be considered for the development of antiviral drug, because of its main role in the viral replication and transcription (151). Among the three groups of genus coronavirus, structures of main proteases (Mpros) from TGEV, HCoV-229E (group I), SARS CoV (group II), and IBV (group III) are available and the substrate-binding pockets present in the cleft between domain I and domain II, and particularly S4, S2, and S1 are greatly conserved among coronaviruses Main proteases (144), (3), (2), (148), and (147).

Coronavirus genome consists of a molecule of single-stranded positive-sense genomic RNA of 27 to 31kb in size with a polyadenylated 3' terminus. Coronaviruses have the largest viral RNA genomes known to date and in case of TGEV it is 28,500 nucleotides in length (3), (2). They are spherically enveloped viruses of about 100 to 160 nm in diameter. Coronavirus genomic RNAs are capped that can serve as mRNAs and the purified genomic RNA can be infectious (80). A leader sequence of 65 to 98 nucleotides is present at the 5' end of the genome and that is also seen at the 5'end of all subgenomic RNAs. Leader sequence is followed by an untranslated region (UTR) of 200 to 400 nucleotides in length. There is another UTR of 200 to 500 nucleotides at the 3' end, followed by poly (A) of variable length **(Figure 2.4)**.

The genome of coronavirus consists of 7 to 14 open reading frames (ORFs) (80). The replicase gene encodes the viral proteins that are essential for TGEV replication and transcription. Two replicative overlapping polyproteins, ORF1a or pp1a (447kDa) and ORF1b or pp1b (754kDa) are encoded by this replicase gene, which are processed by virus-encoded proteases to generate the functional proteins (2). This process of polyprotein processing is mainly achieved by the 33.1 kDa viral cysteine protease, alternatively called 3C-like protease or main protease (Mpro) (3). Other than this Mpro, most of the coronaviruses, with exception to IBV and SARS CoV, encode two papainlike proteases in ORF1a to process the N terminal portion of ORF1a (80). Coronavirus main protease is so called as 3C-like protease, because of the similar substrate specificities between coronavirus Mpro and picornaviral 3C proteases (3Cpro) and the presence of cysteine as the active amino acid in the catalytic site. However, there are structural dissimilarities between the proteases of the two families (2), (151). The replicase polyproteins are cleaved by the Mpro at no less than 11 conserved sites involving Leu-Gln ↓ (Ser, Ala, Gly) sequences ("↓" indicates the cleavage site). This process is initiated by the autoproteolytic cleavage of Mpro from the pp1a and pp1b (45), (26). The crystal structure of coronavirus proteases from TGEV, HCoV-229E, SARS CoV and IBV shows that the molecule comprises of three domains. Domains I and II contain six-stranded antiparallel β barrels and domain III contains a cluster of five helices (3), (144), (2), and (148). The catalytic site of coronaviruses Mpro consists of cysteine and histidine residues, which is conserved in all coronaviruses. The catalytic site of TGEV Mpro is comprised of Cys144 and His41 **(Figure 5.3)**. These proteases show well defined substrate specificities and all known cleavage sites of the proteases contain bulky hydrophobic residues mainly leucine at the P2 position, glutamine at the P1 position, and small aliphatic residues at the P1' position. A unique feature found in coronavirus Main proteases is the presence of a

large C-terminal domain of ~100 amino acid residues that is not found in any other RNA virus 3C-like proteases (3), (2).

To date, there is no inhibitor for TGEV Mpro has been developed other than the existing PRCV attenuated vaccine. Besides this efficacious TGEV vaccine, the development of an antiviral drug with wide range spectrum is very useful, because the developed wide spectrum antiviral agent can be used to control all coronavirus infections. At the same time, the antiviral product is not only useful to control the new virulent forms that evolve from the existing viruses, but also to reduce transmission. Since coronavirus (TGEV) Mpro is essential for the processing of the two immature polyproteins to produce mature or functional viral proteins that are required for viral replication and transcription, the Mpro presents itself as an attractive target for the development of antiviral drugs.

Among many virally encoded components, viral proteases play an important role in the process of virus replication and transcription that are essential for the development of new virions. Viral proteases present themselves as key targets for the development of antivirals. To characterize the proteolytic activity of these viral proteases, a fast, convenient, continuous assay with rapid screening of possible inhibitors is needed.

Earlier, the proteolytic activities of viral proteases were conducted by substrate-analog peptide cleavage assays using conventional HPLC methods (19), (49). These methods are highly time consumable and they require skilled personnel to analyse the data. Hence there is a need for an alternative method that will greatly facilitate large scale screening of potential inhibitors targeting viral proteases without time consuming. One such assay is fluorescence resonance energy transfer (FRET) protease assay, which has been developed and successfully applied for various cellular and viral proteases including severe acute respiratory syndrome coronavirus (SARS CoV), foot-and-mouth disease virus (FMDV), and

transmissible gastroenteritis virus (TGEV) (49), (19), and (147). The results obtained from the FRET based assay are more sensitive and less time-consuming compared with the RP-HPLC method. In FRET based assays, a substrate is designed based on a sequence that is recognized and cleaved by the protease and the protease activity is successfully monitored. In this FRET assay system, substrates are designed in such a way that they have a fluorescence donor on one side of the cleavage site, and a fluorescence quencher on the other side of the cleavage site. EDANS/DABCYL is the commonly used pair, because of the excellent energy overlap of the donor and acceptor (49). In the uncleaved substrate, the donor fluorescence signal is quenched by the nearby fluorescence quencher. Once, the donor and quencher are physically separated by cleavage in the presence of protease, the donor fluorescence is no longer quenched, thereby increasing the fluorescence yield of the donor group. The proteolytic activity can be checked in this way by measuring the increase of fluorescence with time. The cleavage of the substrates can be inhibited by the addition of protease inhibitors to the assay that gives reduced fluorescence intensity **(Figure 2.5)**. This is achieved by the saturation of protease catalytic site with inhibitors. Thus, FRET assay provides rapid screening and identification of potential protease inhibitors. This assay system is also useful in measuring the activity and substrate specificity of a protease.

To date, most of the information available on the proteolytic processing of norovirus proteases is obtained using in vitro transcription-translation assays (38), (121), and (126) and a very few studies are reported on the kinetics of group I norovirus proteases using Fluorogenic (127) and chromogenic substrates (48). However, a fast, convenient, and continuous assay like FRET that would permit the rapid screening of possible protease inhibitors has not yet been reported for proteases from noroviruses GI and GII (150). In the case of picornaviral proteases, FRET based assays have been successfully used to monitor poliovirus 3C protease, foot-and-mouth disease virus 3C protease, and hepatitis A virus 3C

protease (49), (50), (138), and (41). Whereas in the case of coronavirus Mpro, FRET based assays were successfully used for monitoring SARS CoV, TGEV, HCoV-229E, FIPV, MHV, and IBV (19), (12), (53), (146), and (147).

Purpose of the study

The purpose of this study was to characterize the enzymatic activity of related proteases from Norwalk virus, poliovirus, and transmissible gastroenteritis virus by using fluorescence resonance energy transfer (FRET) technique. These proteases share several common characteristics including a typical chymotrypsin-like fold, a Cys residue as a nucleophile in the catalytic triad (or dyad) composed of Cys, His, and Glu (or Asp) residues, and preference for Glu or Gln residue at the P1 position on the substrate. As these proteases cleave the polyprotein of viruses into non-structural functional proteins that are essential for virus replication and production of new virions, the viral proteases presents themselves as an alluring target for the development of antivirals. To conduct such type of research, it is very important to optimize the assay buffer conditions to characterize the enzymatic activity of proteases. So, first our aim is to optimize the assay conditions of the FRET assay for each virus protease and then second, to study the inhibitor profiles of each virus protease using five commercially available standard protease inhibitors (chymostatin, leupeptin, antipain, TPCK, and TLCK).

CHAPTER 3-MATERIALS AND METHODS

The expression and purification of proteases from Norwalk, polio and transmissible gastroenteritis viruses:

A full-length codon-optimized cDNAs corresponding to the total amino acid sequence of Norwalk virus protease (NVpro), poliovirus protease (PVpro), transmissible gastroenteritis virus protease (TGEVpro) were cloned into the pET28a vector (GenScript, Piscataway, NJ). The synthesized cDNA sequences include start and stop codons as well as sequences encoding N-terminal six His amino acid residues for Ni column purification. The plasmid carrying a gene sequence that encodes Norwalk virus protease or poliovirus protease or transmissible gastroenteritis virus protease was transformed into E.*coli* BL21 cells.

Transformation of pET28a vector into the BL21 cells:

Transformation is the process of introducing foreign DNA into the bacterial cells. A volume of 12.5µl thawed BL21 cells were taken in four pre-chilled 1.5 ml eppendorf tubes. To each tube one clone of 5µl pET28a vector carrying amino acid gene sequence for Norwalk virus protease, poliovirus protease, TGEV protease were added and mixed very gently. The cells are incubated on ice for 30 minutes. After 30 minutes of incubation on ice, the cells were subjected to heat shock at 42 ˚C by placing the tubes in a water bath for 45 to 60 seconds. The tubes are immediately transferred on to ice and kept on ice for 2 to 3 minutes. A volume of 250µl pre warmed SOC medium at 37 ˚C was added to the cells and incubated at 37 ˚C for one hour in a shaker incubator.

Small scale protein expression:

Small scale protein expression is useful in testing the cells whether they are producing the protein of interest or not. So to perform small scale protein expression, after one hour of incubation at 37 ˚C, 250µl of transformed cells carrying a recombinant plasmid containing gene sequence that encodes for NV protease, PV protease, TGEV protease were transferred into a 10 mL LB broth. To this 10 ml LB broth a volume of 10µl kanamycin antibiotic (50mg/ml) was added. This LB broth with transformed cells is placed in a shaker incubator at 37 ˚C for an overnight culture. Next morning 5 mL of culture from the tube was replaced with 5 ml of LB broth and 5µl of kanamycin antibiotic and again placed at 37 ˚C in a shaker incubator for one hour incubation. Meanwhile four 250 ml sterile conical glass flasks were taken and 50 ml of LB broth and 50 µl of kanamycin antibiotic (50mg/ml) is added to each flask and placed in a shaker incubator at 37 ˚ C for pre-warming the medium before inoculating with culture. After one hour of incubation, 1 ml of culture was added to the pre-warmed 50 ml medium and optical density readings were noted at 600nm using spectrophotometer. When the optical density of the culture reached 0.4, one ml of uninduced samples were collected in a 1.5 ml eppendorf tubes before inducing with 1mM final concentration of isopropyl β-D-thiogalactopyranosie (IPTG). A volume of 500 µl of IPTG was added to 50 ml of culture medium. The induced cultures are kept at 37 ˚C for 4 hours in a shaker incubator.

SDS-PAGE analysis:

The uninduced samples are centrifuged at 12,000 rpm for one minute and the pellets are collected by discarding the supernatant carefully. The uninduced pellets are carefully resuspended with 250 µl of 1X LDS buffer and sonicated. After 4 hours of induction, the induced cultures are centrifuged at 3000 rpm for 5 minutes. The supernatants were discarded

and the pellets are carefully resuspended with 250 µl of 1X LDS buffer and sonicated. Make sure that sample tubes are on ice while sonication. After sonication, samples were heated at 95 ˚ C for 3 minutes before loading on to the SDS-PAGE gel. A volume of 15 µl of samples was run across a volume of 5 µl EZ-marker ladder. When the samples were fully run, the gel was washed three times with double distilled water for 5 minutes each time and stained with Simply blue for an hour by placing it on shaker. After one hour of staining, the gel was destained with double distilled water by placing it on shaker.

Large scale protein expression:

The process for large scale protein expression is also similar to small scale expression other than the quantity of protein expressed. In large scale protein expression, for overnight culture, 20 ml of LB broth and 20 µl of kanamycin antibiotic (50mg/ml) are taken in a 50 ml tube, inoculated with BL21 cells carrying recombinant plasmid vectors (pET28a) containing genes that code for proteases of interest (Norwalk virus protease or poliovirus 3C protease, or TGEV main protease), and kept in a shaker incubator at 37 ˚ C for overnight. Next day, 10 ml medium from the overnight culture was pipetted out and replaced with10 ml of new LB media and 10 µl of kanamycin (50mg/ml) was added and placed again in shaker incubator at 37 ˚ C for one hour incubation. Meanwhile four large 2 liter sterile glass flasks were taken and to each flask, 200 ml of sterile LB media and 200 µl of kanamycin are added. These flasks were kept in a shaker incubator at 37 ˚ C for pre-warming the media before inoculation with overnight culture. After one hour of incubation, 5 ml of overnight culture was added to each large flask and optical density (OD) readings were measured at 600 nm using spectrophotometer. When the optical density value reached 0.4, each flask was induced with 2 ml of IPTG (1 mM final concentration) and the induced flasks were kept at 37 ˚C for 4 hours.

Sample preparation for purification:

After 4 hours of induction, the cells were harvested by centrifugation at 5000 rpm for 10 minutes. The pellets were collected by carefully discarding the supernatant. To each pellet 5 ml of binding buffer (GE Healthcare His Buffer Kit) was added and the pellets were carefully resuspended. To each 5 ml of resuspended solution 0.035 grams of enzyme lysozyme was added and mixed thoroughly. Now these solutions tubes were placed on ice for 30 minutes and mixed thoroughly for every 5 minutes interval time. After 30 minutes, the samples were sonicated and centrifuged at 10,500 rpm for 20 minutes at 4˚ C. The pellet was discarded and the supernatant was collected. The collected supernatant was filtered using MILLER-HV PVDF 0.45 μm syringe driven filter unit.

Performing purification for the filtered protein sample:

The filtered sample was purified by running across the His GraviTrap affinity column (GE Healthcare). The steps involved in the purification are:

1. First, the bottom tip of the column was cut off, then the top cap was removed, excess liquid from the column was poured off and the column was placed in the workmate column stand.

2. Now the column was equilibrated with 10 ml binding buffer to protect the column from running dry during the run. Allow the binding buffer to completely run through the column.

3. Now the filtered sample was added to the column.

4. Once the sample was completely run, 10 ml of binding buffer was added to the column and allowed it to completely run out of the column.

5. Now 3 ml of elution buffer was added to the column and the eluate was collected in a tube.

Performing dialysis for the eluted protein sample from the column:

To remove the imidazole from the eluted protein sample, dialysis step was performed using dialysis buffer (50 mM sodium phosphate, 100 mM sodium chloride, and 1 mM DTT) at pH-7.2. The process of dialysis was performed using dialysis cassette (Thermo Scientific Slide-A-Lyzer Dialysis Cassette of 10,000 MWCO). The cassette was hydrolysed by placing it in dialysis buffer for 15 to 20 minutes before loading it with the eluted protein sample. A volume of 3 ml eluted protein sample was loaded very gently into the cassette by using 22 $^{1}/_{2}$ gauge syringe needle. Now the loaded cassette was placed in a beaker containing dialysis buffer and magnetic stirrer. The beaker was placed on a magnetic plate. After one hour of dialysis, the buffer in the beaker was replaced. This process was continued for overnight and the buffer was replaced again on the next day. After one or two hours, the sample was collected from the cassette using 22 $^{1}/_{2}$ gauge syringe needle and the concentration of protein in the sample was estimated using Bradford protein assay.

Bradford protein assay:

The Bradford protein assay is a spectroscopic analytical procedure used to measure the concentration of protein in a solution. It was developed by Marion M. Bradford. Here, we used the Bio-Rad Protein Assay to measure the concentration of protein. The Bio-Rad Protein Assay, based on the method of Bradford, is a simple and accurate procedure for determining concentration of solubilized protein. It involves the addition of an acidic dye to protein solution, and subsequent measurement at 595 nm with a spectrophotometer or microplate reader. Comparison to a standard curve provides a relative measurement of protein concentration. The Bio-Rad Protein Assay can also be used with a microplate reader, such as

Bio-Rad's Model 450 and 3550 Microplate Readers. Here, we followed the standard procedure for microtiter plates and it involves the following steps.

1. The dye reagent was prepared by diluting one part of dye reagent concentrate with four parts of double distilled water. This diluted dye reagent may be used for about two weeks when kept at room temperature.

2. Five dilutions of BSA protein standard were prepared. The linear range of protein standard used in this microtiter plate assay is 0.5 mg/ml, 0.4 mg/ml, 0.3mg/ml, 0.2 mg/ml, and 0.1 mg/ml. To obtain these dilutions, first a 1:20 dilution of BSA (10 mg/ml) was prepared by adding 10 µl of BSA to 190 µl of double distilled water. From this 200 µl, a volume of 40, 30, 20, and 10 µl of BSA were added to a volume of 10, 20, 30, and 40 µl of double distilled water respectively to get the above mentioned linear range of protein standards.

3. To prepare unknown protein sample, first a 1:4 dilution of unknown protein sample (25 µl of protein sample + 75 µl of double distilled water) was prepared, which was further serially diluted in double distilled water by a twofold dilution factor up to 1:128 dilution.

4. 10 µl each of standard and unknown protein samples were added into separate microtiter plate wells. All standard and unknown protein samples were assayed in triplicates.

5. A volume of 200 µl of diluted dye reagent is added to each well and mixed thoroughly. The plate was incubated at room temperature for 3 to 5 minutes and absorbance was measured at 595 nm using a microplate reader.

6. The concentration of protein in the sample was calculated by comparing its absorbance value with the standard curve that is plotted using Microsoft excel by plotting protein concentration as the dependent variable on the Y-axis and absorbance as the independent variable on the X-axis.

SDS-PAGE analysis of dialysed protein samples:

After dialysis, 10 µl of each protein sample from Norwalk virus protease, poliovirus 3C protease, and transmissible gastroenteritis virus main protease was mixed with 10 µl of 2X LDS buffer respectively and the samples were heated at 95 ˚ C for 3 minutes before loading on to the SDS-PAGE gel. A volume of 15 µl of samples was run across a volume of 5 µl EZ-marker ladder. When the samples were fully run, the gel was washed three times with double distilled water for 5 minutes each time and stained with Simply blue for an hour by placing it on shaker. After one hour of staining, the gel was destained with double distilled water by placing it on shaker.

FRET enzymatic activity characterization of proteases from Norwalk virus, poliovirus, and transmissible gastroenteritis virus

Substrates:

Substrate for Norwalk virus protease:

Two Fluorogenic substrates with edans and dabcyl pair were tested as substrates for Norwalk virus protease enzymatic activity. The 14-amino acid fluorogenic substrate edans-EPDFHLQGPEDLAK-dabcyl and 7-amino acid fluorogenic substrate edans-DFHLQGP-dabcyl derived from the P7-P7’ and P5-P2’ amino acid residues on the NS2/NS3 cleavage site in ORF1 of Norovirus, respectively, were synthesized (GenScript, Piscataway, NJ). The representation of substrate residues for P1 and P1’ begins at the scissile bond and counting

towards the N-terminus is called P1 side and towards the C-terminus is called P1' side, respectively as suggested by Schechter and Berger [12].

Substrate for poliovirus 3C protease:

The 9-amino acid fluorogenic substrate edans-EALFQGPLQ-dabcyl was synthesized (GenScript, Piscataway, NJ) and tested with poliovirus 3C protease and the designation of substrate amino acid residues for P1 and P1' begins at the scissile bond and counting towards the N-terminus is called P1 side and towards the C-terminus is called P1' side, respectively as suggested by Schechter and Berger [12].

Substrate for transmissible gastroenteritis virus 3C-like or main protease:

The 12-amino acid fluorogenic substrate edans-VNSTLQSGLRKM-dabcyl was synthesized (GenScript, Piscataway, NJ) and tested for transmissible gastroenteritis virus main protease.

FRET protease assay for Norwalk virus protease, poliovirus 3C protease, and transmissible gastroenteritis virus main protease to determine the concentration of protease that shows maximum enzymatic activity:

Stock solutions (10 mM) of all the three protease substrates were prepared by dissolving in DMSO and these substrates were used for all reactions.

Optimization of Norwalk virus protease concentration:

The 25 µl of assay buffer containing different concentrations of Norwalk virus protease (0.5, 1.0, 3.0, 5.0, and 10 µM,) are mixed with 25 µl of assay buffer containing same concentration (20 µM) of 7-amino acid fluorogenic substrate (edans-DFHLQGP-dabcyl) in a 96-well black plate and incubated up to 30 min at 37 ˚C. The assay buffer consists of 20 mM HEPES, 120

mM NaCl, 0.4 mM EDTA, 4 mM DTT and 20 % Glycerol. The pH of the buffer is adjusted to 8.0. The fluorescence signals were detected using an excitation wavelength of 360 nm and an emission wavelength of 460 nm on a fluorescence microplate reader (FLx800, Biotek, Winooski, VT) every 5 min for 30 min. The fluorescence readings from the substrate control well (background) were subtracted from the readings of the other wells to get the relative fluorescence unit (RFU).

Optimization of poliovirus 3C protease concentration:

The 25 μl of assay buffer containing different concentrations of poliovirus protease (1.0, 5.0, 10, and 15 μM,) are mixed with 25 μl of assay buffer containing same concentration (20 μM) of fluorogenic substrate (edans-EALFQGPLQ-dabcyl) in a 96-well black plate and incubated up to 30 min at 37 ˚C. The assay buffer consists of 20 mM HEPES, 120 mM NaCl, 0.4 mM EDTA, 4 mM DTT and 20 % Glycerol. The pH of the buffer is adjusted to 6.0. The fluorescence signals were detected using an excitation wavelength of 360 nm and an emission wavelength of 460 nm on a fluorescence microplate reader (FLx800, Biotek, Winooski, VT) every 5 min for 30 min. The fluorescence readings from the substrate control well (background) were subtracted from the readings of the other wells to get the relative fluorescence unit (RFU).

Optimization of transmissible gastroenteritis virus main protease concentration:

The 25 μl of assay buffer containing different concentrations of transmissible gastroenteritis virus main protease (0.1, 0.5, 1.0, and 3.0 μM) are mixed with 25 μl of assay buffer containing same concentration (20 μM) of fluorogenic substrate (edans-VNSTLQSGLRKM) in a 96-well black plate and incubated up to 30 min at 37 ˚C. The assay buffer consists of 20 mM HEPES, 120 mM NaCl, 0.4 mM EDTA, 4 mM DTT and 20 % Glycerol. The pH of the buffer is adjusted to 6.0. The fluorescence signals were detected using an excitation

wavelength of 360 nm and an emission wavelength of 460 nm on a fluorescence microplate reader (FLx800, Biotek, Winooski, VT) every 5 min for 30 min. The fluorescence readings from the substrate control well (background) were subtracted from the readings of the other wells to get the relative fluorescence unit (RFU).

FRET protease assay for the optimization of assay buffer conditions for proteases from Norwalk virus protease, poliovirus 3C protease, and transmissible gastroenteritis virus main protease:

Optimization of glycerol percentage:

Optimization of Glycerol is very important for assay buffer, because it provides stability to the enzyme by preventing the rapid loss of enzyme activity. Different percentages of Glycerol ranging from 20%, 30%, 40%, and 50% are tested by keeping the remaining all other conditions of the buffer (20 mM HEPES, 120 mM NaCl, 0.4 mM EDTA, 4 mM DTT) as constant. In case of Norwalk virus protease, pH is adjusted to 8.0, whereas for poliovirus and transmissible gastroenteritis virus proteases, pH is adjusted to 6.0.

Optimization of glycerol percentage for Norwalk virus protease:

The 25 µl of assay buffer containing 5 µM concentration of Norwalk virus protease is mixed with 25 µl of assay buffer containing 10 µM concentration of 7-amino acid fluorogenic substrate (edans-DFHLQGP-dabcyl) in a 96-well black plate and incubated up to 30 min at 37 ˚C. The fluorescence signals were detected using an excitation wavelength of 360 nm and an emission wavelength of 460 nm on a fluorescence microplate reader (FLx800, Biotek, Winooski, VT) every 5 min for 30 min. The fluorescence readings from the substrate control

well (background) were subtracted from the readings of the other wells to get the relative fluorescence unit (RFU).

Optimization of glycerol percentage for poliovirus 3C protease:

The 25 µl of assay buffer containing 5 µM concentration of poliovirus protease is mixed with 25 µl of assay buffer containing 10 µM concentration of fluorogenic substrate (edans-EALFQGPLQ-dabcyl) in a 96-well black plate and incubated up to 30 min at 37 ˚C. The fluorescence signals were detected using an excitation wavelength of 360 nm and an emission wavelength of 460 nm on a fluorescence microplate reader (FLx800, Biotek, Winooski, VT) every 5 min for 30 min. The fluorescence readings from the substrate control well (background) were subtracted from the readings of the other wells to get the relative fluorescence unit (RFU).

Optimization of glycerol percentage for transmissible gastroenteritis virus main protease:

The 25 µl of assay buffer containing 2 µM concentration of transmissible gastroenteritis virus main protease is mixed with 25 µl of assay buffer containing 10 µM concentration of fluorogenic substrate (edans-VNSTLQSGLRKM) in a 96-well black plate and incubated up to 30 min at 37 ˚C. The fluorescence signals were detected using an excitation wavelength of 360 nm and an emission wavelength of 460 nm on a fluorescence microplate reader (FLx800, Biotek, Winooski, VT) every 5 min for 30 min. The fluorescence readings from the substrate control well (background) were subtracted from the readings of the other wells to get the relative fluorescence unit (RFU).

Optimization of dithiothreitol (DTT) concentration:

Most proteins contain free "SH" groups. Upon oxidation, "SH" turn from intra to inter molecular S-S bonds, which usually results in the loss of enzyme activity. A wide variety of agents are available to prevent the disulphide bond formation. One among them is DTT, a reducing agent which is helpful in maintaining the protein in the reduced state by preventing oxidation of cysteines. Different concentrations of DTT (Dithiothreitol) from 0 mM, 2.0 mM, 4.0 mM, 6.0 mM, 8.0 mM, 10 mM and 30 mM range were tested for Norwalk virus protease , transmissible gastroenteritis virus protease and Poliovirus 3C protease. Other than DTT concentration, the assay buffer for Norwalk virus consists of 20 mM HEPES, 120 mM NaCl, 0.4 mM EDTA, and 50 % Glycerol and the pH is 8.0. For poliovirus and transmissible gastroenteritis virus proteases, the assay buffer consists of 20 mM HEPES, 120 mM NaCl, 0.4 mM EDTA, and 20 % Glycerol and the pH is adjusted to 6.0.

Optimization of DTT concentration for Norwalk virus protease:

The 25 µl of assay buffer containing 5 µM concentration of Norwalk virus protease is mixed with 25 µl of assay buffer containing 10 µM concentration of the 7-amino acid fluorogenic substrate (edans-DFHLQGP-dabcyl) in a 96-well black plate and incubated up to 30 min at 37 ˚C. The fluorescence signals were detected using an excitation wavelength of 360 nm and an emission wavelength of 460 nm on a fluorescence microplate reader (FLx800, Biotek, Winooski, VT) every 5 min for 30 min. The fluorescence readings from the substrate control well (background) were subtracted from the readings of the other wells to get the relative fluorescence unit (RFU).

Optimization of DTT concentration for poliovirus 3C protease:

The 25 µl of assay buffer containing 5 µM concentration of poliovirus protease is mixed with 25 µl of assay buffer containing 10 µM concentration of fluorogenic substrate (edans-EALFQGPLQ-dabcyl) in a 96-well black plate and incubated up to 30 min at 37 ˚C. The fluorescence signals were detected using an excitation wavelength of 360 nm and an emission wavelength of 460 nm on a fluorescence microplate reader (FLx800, Biotek, Winooski, VT) every 5 min for 30 min. The fluorescence readings from the substrate control well (background) were subtracted from the readings of the other wells to get the relative fluorescence unit (RFU).

Optimization of DTT concentration for transmissible gastroenteritis virus main protease:

The 25 µl of assay buffer containing 2 µM concentration of transmissible gastroenteritis virus main protease is mixed with 25 µl of assay buffer containing 10 µM concentration of fluorogenic substrate (edans-VNSTLQSGLRKM) in a 96-well black plate and incubated up to 30 min at 37 ˚C. The fluorescence signals were detected using an excitation wavelength of 360 nm and an emission wavelength of 460 nm on a fluorescence microplate reader (FLx800, Biotek, Winooski, VT) every 5 min for 30 min. The fluorescence readings from the substrate control well (background) were subtracted from the readings of the other wells to get the relative fluorescence unit (RFU).

Optimization of NaCl concentration:

Some enzymes require a polar medium with high ionic strength to maintain full activity. Low ionic strength favours the rapid inactivation of the enzyme. For these infrequent conditions, KCl or NaCl may be used to raise the ionic strength of the solution, thereby providing the

stability to the enzyme Different concentrations of NaCl i.e., 0 mM, 50 mM, 100 mM, 200 mM, and 400 mM were tested for Norwalk virus, poliovirus, and transmissible gastroenteritis virus proteases. The remaining buffer conditions for Norwalk virus protease are 20 mM HEPES, 0 mM DTT, 0.4 mM EDTA, 50 % glycerol, and pH is adjusted to 8.0. The remaining assay buffer conditions for poliovirus and transmissible gastroenteritis virus proteases are 20 mM HEPES, 0 mM DTT, 0.4 mM EDTA, 20 % glycerol, and pH is adjusted to 6.0.

Optimization of NaCl concentration for Norwalk virus protease:

The 25 µl of assay buffer containing 5 µM concentration of Norwalk virus protease is mixed with 25 µl of assay buffer containing 10 µM concentration of the 7-amino acid fluorogenic substrate (edans-DFHLQGP-dabcyl) in a 96-well black plate and incubated up to 30 min at 37 ˚C. The fluorescence signals were detected using an excitation wavelength of 360 nm and an emission wavelength of 460 nm on a fluorescence microplate reader (FLx800, Biotek, Winooski, VT) every 5 min for 30 min. The fluorescence readings from the substrate control well (background) were subtracted from the readings of the other wells to get the relative fluorescence unit (RFU).

Optimization of NaCl concentration for poliovirus 3C protease:

The 25 µl of assay buffer containing 5 µM concentration of poliovirus protease is mixed with 25 µl of assay buffer containing 10 µM concentration of fluorogenic substrate (edans-EALFQGPLQ-dabcyl) in a 96-well black plate and incubated up to 30 min at 37 ˚C. The fluorescence signals were detected using an excitation wavelength of 360 nm and an emission wavelength of 460 nm on a fluorescence microplate reader (FLx800, Biotek, Winooski, VT) every 5 min for 30 min. The fluorescence readings from the substrate control well

(background) were subtracted from the readings of the other wells to get the relative fluorescence unit (RFU).

Optimization of NaCl concentration for transmissible gastroenteritis virus main protease:

The 25 µl of assay buffer containing 2 µM concentration of transmissible gastroenteritis virus main protease is mixed with 25 µl of assay buffer containing 10 µM concentration of fluorogenic substrate (edans-VNSTLQSGLRKM) in a 96-well black plate and incubated up to 30 min at 37 ˚C. The fluorescence signals were detected using an excitation wavelength of 360 nm and an emission wavelength of 460 nm on a fluorescence microplate reader (FLx800, Biotek, Winooski, VT) every 5 min for 30 min. The fluorescence readings from the substrate control well (background) were subtracted from the readings of the other wells to get the relative fluorescence unit (RFU).

Optimization of pH range:

Biochemical reactions are especially sensitive to pH. Many chemical reactions are affected by the acidity of the solution in which they occur. In order for a particular reaction to occur at an appropriate rate, the pH of the reaction medium must be controlled. Different levels of pH ranging from 6, 7, 8, 9 and 10 were tested for all the three viral proteases. Other than pH and glycerol percentage (50 % for Norwalk virus protease, 20 % for poliovirus and transmissible gastroenteritis virus proteases), the remaining assay buffer conditions (20 mM HEPES, 0 mM DTT, 0.4 mM EDTA, and 0 mM NaCl) are similar for all the three viral proteases.

Optimization of pH for Norwalk virus protease:

The 25 µl of assay buffer containing 5 µM concentration of Norwalk virus protease is mixed with 25 µl of assay buffer containing 10 µM concentration of the 7-amino acid fluorogenic

substrate (edans-DFHLQGP-dabcyl) in a 96-well black plate and incubated up to 30 min at 37 ˚C. The fluorescence signals were detected using an excitation wavelength of 360 nm and an emission wavelength of 460 nm on a fluorescence microplate reader (FLx800, Biotek, Winooski, VT) every 5 min for 30 min. The fluorescence readings from the substrate control well (background) were subtracted from the readings of the other wells to get the relative fluorescence unit (RFU).

Optimization of pH for poliovirus protease:

The 25 µl of assay buffer containing 5 µM concentration of poliovirus protease is mixed with 25 µl of assay buffer containing 10 µM concentration of fluorogenic substrate (edans-EALFQGPLQ-dabcyl) in a 96-well black plate and incubated up to 30 min at 37 ˚C. The fluorescence signals were detected using an excitation wavelength of 360 nm and an emission wavelength of 460 nm on a fluorescence microplate reader (FLx800, Biotek, Winooski, VT) every 5 min for 30 min. The fluorescence readings from the substrate control well (background) were subtracted from the readings of the other wells to get the relative fluorescence unit (RFU).

Optimization of pH for transmissible gastroenteritis virus main protease:

The 25 µl of assay buffer containing 2 µM concentration of transmissible gastroenteritis virus main protease is mixed with 25 µl of assay buffer containing 10 µM concentration of fluorogenic substrate (edans-VNSTLQSGLRKM-dabcyl) in a 96-well black plate and incubated up to 30 min at 37 ˚C. The fluorescence signals were detected using an excitation wavelength of 360 nm and an emission wavelength of 460 nm on a fluorescence microplate reader (FLx800, Biotek, Winooski, VT) every 5 min for 30 min. The fluorescence readings from the substrate control well (background) were subtracted from the readings of the other wells to get the relative fluorescence unit (RFU).

FRET protease assay with protease inhibitors for proteases from Norwalk virus protease, poliovirus 3C protease, and transmissible gastroenteritis virus main protease:

To perform FRET protease assay with protease inhibitors, a variety of commercially available standard protease inhibitors including serine protease inhibitors such as chymostatin, leupeptin, N-tosyl-L-phenylalanine chloromethyl ketone (TPCK), and tosyl-L-lysine chloromethyl ketone (TLCK) and a papain-like cysteine inhibitor were obtained from Sigma-Aldrich (St Louis, MO). Stock solutions of these inhibitors (10 mM) were prepared by dissolving in DMSO. Other than pH and glycerol percentage (pH-8, glycerol-50 % for Norwalk virus protease; pH-6, glycerol-20 % for poliovirus and transmissible gastroenteritis virus proteases), the remaining assay buffer conditions (20 mM HEPES, 0 mM DTT, 0.4 mM EDTA, and 0 mM NaCl) are similar for all the three viral proteases.

Rapid screening of protease inhibitors for Norwalk virus protease:

For rapid screening of protease inhibitors, 5 μM of Norwalk virus protease and 10 μM of substrate that gave strong signal were used. All the five inhibitors (chymostatin, leupeptin, TPCK, TLCK, and antipain) at a final concentration of 50 μM and 100 μM were pre-incubated with 5 μM Norwalk virus protease in 25 μl of assay buffer for 30 minutes at 37˚ C. After 30 minutes of pre-incubation, 25 μl of assay buffer containing 10 μM of substrate (edans-DFHLQGP-dabcyl) was added to a 96-well black plate and incubated up to 60 minutes at 37 ˚C. After 60 minutes of incubation, the fluorescence signals were detected using an excitation wavelength of 360 nm and an emission wavelength of 460 nm on a fluorescence microplate reader (FLx800, Biotek, Winooski, VT). RFU was calculated by subtracting the fluorescent readings of substrate control well (background) from the readings of the other wells. The reduction of cleaved products by each preparation was calculated by the comparison to that of control well without any inhibitor.

Rapid screening of protease inhibitors for poliovirus 3C protease:

Fast screening of protease inhibitors such as chymostatin, leupeptin, TPCK, TLCK, and antipain for poliovirus 3C protease was performed by using 5 μM of protease concentration and 10 μM of (edans-EALFQGPLQ-dabcyl) substrate concentration that gave good signal. Inhibitors at a concentration of 50 μM and 100 μM are used and the procedure for rapid screening of inhibitors is similar to that of Norwalk virus protease.

Rapid screening of protease inhibitors for transmissible gastroenteritis virus main protease:

A concentration of 2 μM protease and 10 μM substrate concentrations (edans-VNSTLQSGLRKM-dabcyl) that gave good signal in the studies were used for the quick screening of protease inhibitors including chymostatin, leupeptin, TPCK, TLCK, and antipain for transmissible gastroenteritis virus main protease. Inhibitors at a concentration of 50 μM and 100 μM were tested and the procedure followed is similar to that of rapid screening of Norwalk virus protease inhibitors.

Determination of IC50 of chymostatin and TPCK

IC50 is the concentration of the inhibitor that inhibited half of the activity of enzyme. The IC50 of an inhibitor can be determined by constructing a dose-response curve.

Determination of IC50 of chymostatin for Norwalk virus protease:

Among the five inhibitors tested, Norwalk virus protease was strongly inhibited by chymostatin at a concentration of 50 μM. So to evaluate the detailed efficacy of chymostatin, continuous monitoring of activity of Norwalk virus protease was performed. Protease (5 μM, concentration) was mixed with serially diluted chymostatin (0~50 μM) in a 25 μl of assay buffer and incubated at 37˚ C for 30 minutes, followed by the addition of 25 μl of assay

buffer containing 20 μM substrate (edans-DFHLQGP-dabcyl). The mixture was incubated at 37˚C for 60 minutes, and fluorescent readings were converted to the concentration of the cleavage product as described previously in Materials and Methods, and the dose-dependent FRET inhibition curves were fitted with variable slope using WinSAT software in order to determine the concentration of chymostatin that reduces enzyme velocity by half (IC_{50}) value.

Determination of IC50 of TPCK for poliovirus 3C protease:

Among the five inhibitors tested, poliovirus protease was strongly inhibited by TPCK at a concentration of 100 μM compared to 50 μM. So to evaluate the detailed efficacy of TPCK, continuous monitoring of activity of Norwalk virus protease was performed. Protease (5 μM, concentration) was mixed with serially diluted TPCK (0~150 μM) in a 25 μl of assay buffer and incubated at 37˚ C for 30 minutes, followed by the addition of 25 μl of assay buffer containing 20 μM substrate (edans- EALFQGPLQ -dabcyl). The mixture was incubated at 37˚C for 60 minutes, and fluorescent readings were converted to the concentration of the cleavage product as described previously in Materials and Methods, and the dose-dependent FRET inhibition curves were fitted with variable slope using WinSAT software in order to determine the concentration of chymostatin that reduces enzyme velocity by half (IC50) value.

CHAPTER 4-RESULTS

The protein expression and SDS-PAGE analysis

The SDS-PAGE separates the proteins present in each sample fraction into separate bands and it can be used to estimate the molecular weight of each protein based on its migration distance. The results from the SDS-PAGE analysis for small scale protein expression confirmed that the clones that are transferred into BL21 cells are expressing the proteases of interest. These results are strongly supported by the results from the SDS-PAGE analysis of purified or dialysed samples that showed strong bands with their respective molecular weight. The results from the SDS-PAGE analysis showed a predicted molecular weight of 19 kDa for Norwalk virus protease, 20 kDa for poliovirus protease, and 33 kDa for transmissible gastroenteritis virus (**Figure 4.1**).

Estimation of protein concentration in the samples:

The concentrations of purified protein samples were estimated by using Bio-Rad microtiter plate (Bradford) assay. A concentration of 6 mg/ml of Norwalk virus protease was expressed. For poliovirus 3C protease, the concentration of protein estimated in the purified sample was 7 mg/ml, and for transmissible gastroenteritis virus main protease the protein concentration was estimated as 3 mg/ml.

Optimization of protease concentrations:

All the fluorogenic FRET substrates were efficiently cleaved by their respective proteases.

Norwalk virus protease:

Norwalk virus protease was tested with two substrates, one is 7-amino acids length (edans-DFHLQGP-dabcyl) and the other is 14-amino acids length (edans-EPDFHLQGPEDLAK-dabcyl). Between the two substrates tested for Norwalk virus protease, substrate with 7-amino acids (edans-DFHLQGP-dabcyl) was more efficiently cleaved and this substrate is used for the entire study for Norwalk virus protease. Among the different concentrations of protease tested with same concentration of the substrate for Norwalk virus, there is an increase in the activity of enzyme with increase in protease concentration **(Figure 4.2.A)**. We used a protease concentration of 5μM and a substrate concentration of 10μM for further studies of Norwalk virus protease.

Poliovirus protease:

The substrate, edans-EALFQGPLQ-dabcyl was used for the enzymatic activity characterization of poliovirus 3C protease. Increase in the relative fluorescence confirmed that poliovirus 3C protease cleaved the substrate very effectively. Hence, this FRET substrate was used for the entire poliovirus enzymatic activity characterization and inhibition studies. Poliovirus protease was also tested with different concentrations of protease and same concentration of substrate. Relative fluorescence increased with the increase in protease concentration. This suggests that poliovirus protease also showed similar kind of activity like Norwalk virus protease i.e. increase in the activity of enzyme with the increase in the concentration of protease **(Figure 4.2.B)**. We used a protease concentration of 5μM and a substrate concentration of 10μM for further studies of poliovirus 3C protease.

Transmissible gastroenteritis virus main protease:

TGEV main protease was tested with the 12-amino acid substrate edans-VNSTLQSGLRKM-dabcyl for enzymatic activity characterization. A substrate with similar amino acid sequence is used for SARS-CoV main protease by Chen. S. et. al (2005). The TGEV main protease cleaved the substrate efficiently. Hence, this substrate was used for the enzymatic activity characterization and inhibition studies of TGEV main protease. Different concentrations of protease were tested with same concentration of substrate. Transmissible gastroenteritis virus main protease also showed increase in enzyme activity with the increase in the concentration of protease **(Figure 4.2.C)**. We used a protease concentration of 2μM and a substrate concentration of 10μM for further studies of transmissible gastroenteritis virus main protease.

Optimization of assay buffer conditions:

Assay buffer optimization is very important for the study of enzymatic activity characterization, because different enzymes need different buffer conditions. Assay buffer conditions were optimized for the each viral protease by testing with different percentages of glycerol, different concentrations of NaCl, DTT, and different pH values.

Glycerol percentage:

All three viral proteases were tested with different percentages of glycerol ranging from 20% to 50% in the assay buffer.

Norwalk Virus Protease: For the different percentages of glycerol tested, there is an increase in the activity of Norwalk virus protease with the increase in glycerol percentage. A maximum enzymatic activity was observed at 50 % glycerol for Norwalk virus protease. Previously, during preliminary studies, Norwalk virus protease showed maximum activity at 70% glycerol also. But it became hard to mix the substrate at that percentage of glycerol in

the assay buffer. Hence, we selected 50% glycerol in buffer because at this percentage also, it showed maximum activity **(Figure 4.3.A)**.

Poliovirus Protease: Among the different percentages of glycerol tested, Poliovirus 3C protease showed maximum activity at 20 and 30 % glycerol in the assay buffer and there is a decrease in the activity with the increase of glycerol percentage for poliovirus protease. Taking the above results into consideration, we selected 20% of glycerol in the assay buffer for further studies of poliovirus protease **(Figure 4.3.B)**.

Transmissible gastroenteritis virus main protease:

Transmissible gastroenteritis virus main protease also showed similar kind of activity like poliovirus protease at all glycerol percentages tested. TGEV main protease was also tested with different percentages of glycerol in the buffer. It showed maximum activity at 20% and 30% of glycerol in the assay buffer. Further increase in the percentage of glycerol in the assay buffer reduced the enzymatic activity of TGEV main protease. Hence, we used 20% of glycerol for the further studies of TGEV main protease **(Figure 4.3.C)**.

DTT concentration:

Different concentrations of DTT ranging from 0 mM to 30 mM were tested for all the three viral proteases.

Norwalk virus protease:

Between the ranges of 0 mM to 10 mM the activity of the Norwalk virus protease is almost similar and its activity decreased with the increase in the concentration of DTT (30 mM) in the assay buffer. In later studies of Norwalk virus protease, we used buffer without DTT in it **(Figure 4.4.A)**.

Poliovirus Protease:

Poliovirus protease also showed the same kind of results that are similar to Norwalk virus protease. From 0 mM to 10 mM it showed similar kind of activity and it showed reduced activity as like Norwalk virus protease with the increase in the concentration of DTT (30 mM) in the assay buffer. All further studies of poliovirus 3C protease were carried without DTT in assay buffer **(Figure 4.4.B)**.

Transmissible gastroenteritis virus main protease:

Transmissible gastroenteritis virus showed activity that is different from Norwalk virus protease and poliovirus 3C protease. Increase in the concentration of DTT up to 30 mM doesn't show any kind of effect on the activity of TGEV main protease. Transmissible gastroenteritis virus main protease showed similar kind of activity (positive activity) from 0 mM to 30 mM DTT **(Figure 4.4.C)**. Therefore, remaining all studies of TGEV main protease was carried without DTT in the assay buffer.

NaCl concentration:

Different concentrations of NaCl from 0 mM to 400 mM in assay buffer were tested for each protease.

Norwalk virus protease:

Norwalk virus protease showed maximum activity without NaCl in buffer, or at low concentrations up to 50 mM. There is a decrease in the activity of Norwalk virus protease with further increase in the concentration of NaCl **(Figure 4.5.A)**. Hence, assay buffer without NaCl was used for Norwalk virus protease.

Poliovirus 3C protease:

Poliovirus 3C protease behaved differently from the other two proteases. It showed almost similar activity at all concentrations of NaCl tested. As a result, we selected assay buffer without NaCl for poliovirus 3C protease studies **(Figure 4.5.B)**.

Transmissible gastroenteritis virus main protease:

Increased concentrations of NaCl showed reduced activity of TGEV main protease. Transmissible gastroenteritis virus main protease showed maximum activity without NaCl in the assay buffer. Increase in the concentration of NaCl decreased the activity of transmissible gastroenteritis virus protease **(Figure 4.5.C)**. Further studies of TGEV main protease was also carried without NaCl in the assay buffer like Norwalk virus protease and poliovirus 3C protease.

pH range:

All the three viral proteases; Norwalk virus protease, poliovirus 3C protease, and TGEV main protease were tested at different pH points from 6.0 to 10.

Norwalk virus protease:

At various pH values tested, the Norwalk virus protease exhibited a stable proteolytic activity between pH 7.0 – 9.0 and displayed less activity at pH 6.0 and 10. Between pH points 7.0 and 9.0, Norwalk virus protease showed maximum activity at pH-8.0 **(Figure 4.6.A)**. Hence, assay buffer with pH 8.0 was used for Norwalk virus protease studies.

Poliovirus 3C protease:

Among various pH points tested, poliovirus 3C protease exhibited maximum proteolytic activity between pH 6.0 to 8.0, and displayed very less activity at pH 9.0 and 10. Between pH

values 6.0 and 8.0, poliovirus exhibited maximum activity at pH 6.0 **(Figure 4.6.B)**. So, assay buffer with pH 6.0 was used for further studies of poliovirus 3C protease.

Transmissible gastroenteritis virus main protease:

Among the different pH values tested, transmissible gastroenteritis virus displayed maximum proteolytic activity at pH 6.0, and showed medium activity at pH 7.0 and 8.0, but very less activity at the pH 9.0 and 10.0 **(Figure 4.6.C)**. Assay buffer with pH 6.0 was selected to use for further studies of transmissible gastroenteritis virus main protease.

Rapid screening of protease inhibitors:

Five commercially available protease inhibitors including serine protease inhibitors such as chymostatin, leupeptin, TPCK, and TLCK and a papain-like cysteine protease inhibitor (antipain) were used to identify the inhibitor specificity of each protease. All inhibitors were tested at 50 μM and 100 μM concentrations.

Norwalk virus protease:

Among all the five protease inhibitors (50 μM and 100 μM) tested, Norwalk virus protease was inhibited strongly by chymostatin at both concentrations and a lesser extent by TPCK **(Figure 4.7.A)**. As Norwalk virus protease was inhibited by more than 50% at 50 μM concentration of chymostatin, IC_{50} curve was calculated by taking the 50 μM concentration of chymostatin as maximum concentration.

Poliovirus 3C protease:

Poliovirus 3C protease was inhibited by TPCK and remaining all four protease inhibitors were failed to inhibit the protease. Poliovirus 3C protease was strongly inhibited by 100 μM concentration of TPCK than 50 μM concentration **(Figure 4.7.B)**.

Transmissible gastroenteritis virus main protease:

Transmissible gastroenteritis virus main protease showed less inhibition profile with all the protease inhibitors tested. Interestingly, all these protease inhibitors inhibited the transmissible gastroenteritis virus up to a small extent. Among the five inhibitors, it was slightly more inhibited by TPCK similar at 50 and 100 μM **(Figure 4.7.C)**.

Dose response curve analysis:

Commercially available protease inhibitors including serine protease inhibitors such as chymostatin, leupeptin, TPCK, and TLCK and a papain-like cysteine protease inhibitor (antipain) were used to identify the inhibitor specificity of each protease.

Norwalk virus protease:

Norwalk virus protease was strongly inhibited by only chymostatin at both concentrations tested for the protease (50 μM and 100 μM) .Chymostatin at 50 μM significantly (more than 50 %) decreased the fluorescence intensity from inhibition of substrate cleavage by Norwalk virus protease. However, antipain, TLCK, TLCK, and leupeptin failed to inhibit Norwalk virus protease. Thus, IC_{50} of chymostatin was determined for Norwalk virus protease. The dose response curve for chymostatin on Norwalk virus protease was obtained by using WinSTAT software and the IC_{50} of chymostatin for Norwalk virus protease was determined as 37.321 μM chymostatin **(Figure 4.8.A)**.

Poliovirus 3C protease:

TPCK at 100 μM significantly inhibited the poliovirus 3C protease. The remaining four protease inhibitors chymostatin, TLCK, leupeptin and antipain failed to inhibit poliovirus 3C

protease. Thus, IC_{50} of TPCK (150 μM) was determined for poliovirus 3C protease. The dose response curve for TPCK on poliovirus 3C protease were obtained by using WinSTAT software and the IC_{50} of TPCK on poliovirus 3C protease is determined as 68.93 μM of TPCK **(Figure 4.8.B)**.

CHAPTER 5- DISCUSSION AND CONCLUSIONS

Proteases are present in all animal cells and in various subcellular components. An enormous variety of biochemical reactions that comprise life are almost mediated by a remarkable array of biological catalysts called as enzymes. Proteases also known as proteolytic enzymes or proteinases refer to a group of enzymes whose catalytic function is to hydrolyze peptide bonds in proteins. They account for about 2% of the human genome, 1-5% of genomes of disease causing infectious organisms like bacteria, fungi, viruses, and other forms of life (43), (1). Proteolytic enzymes are essential for most physiological process, but their overexpression or unregulated actions can lead to many chronically debilitating diseases associated with, for example, the CNS and cardiovascular systems, inflammatory and neurodegenerative disease conditions, and viral and parasitic infections.

Viruses are main pathogenic organisms that can cause a large number of serious diseases in humans. For example HIV is responsible for AIDS; herpes simplex virus causes oral lesions and genital herpes, hepatitis C virus causes chronic liver diseases cirrhosis and hepatocellular carcinoma. There are roughly 170 million chronic HCV carriers in the world, or roughly about 3% of the world population (132). As most of the viruses contain RNA as their genetic material, the chances of mutation rate is very high and different new strains are evolving from the existing strains. Because of their clinical and scientific importance viruses have been under intensive study.

The discovery and development of new antiviral compounds is an important part of the research on viruses. One of the viral components that can be targeted for the development of antiviral strategies is viral encoded proteases that are essential for the survival of viruses, as well as their infectivity and virulence. These proteases can be targeted for the development of antiviral agents because viral proteases play an important role in the processing of viral

precursor polyproteins essential to produce functional viral proteins that are required for transcription and translation of viral proteins. In the face of such very strong evidence for proteases being important mediators of viral pathogenesis, it is not surprising that there has been broad research activity and materials targeted to the development of potent and selective protease inhibitors, particularly over the last 20 years.

In this study, we tried to characterize the enzymatic activity of proteases from Norwalk virus, poliovirus, and transmissible gastroenteritis virus by using fast and continuous FRET assay and performed the rapid screening of protease inhibitors to figure out potential protease inhibitors. These proteases share many common characteristics including a typical chymotrypsin-like fold, a Cys residue as a nucleophile in the catalytic triad (or dyad) **(Figures 5.1, 5.2, and 5.3)** composed of Cys, His and Glu (or Asp) residues, and a preference for an Glu or Gln residue at the P1 position on the substrate.

Noroviruses are recognized as the leading cause of food born-disease outbreaks causing acute gastroenteritis around the worldwide, and in the United States (72). They spread from person to person, through contaminated food or water, and by touching contaminated surfaces. Norovirus infections are mainly confined to schools, hospitals, cruise ships, hotels, and military troops and they can affect people of all ages (60). These are considered as highly contagious because very few particles less than 100 can initiate infection in a healthy individual (21). The symptoms of acute gastroenteritis are diarrhea, vomiting, and stomach pain. Long-lasting immunity following natural infection is very rare. The genetic variability in circulating noroviruses indicates that individuals are likely to be repeatedly infected during their life time (77). Currently there is no vaccine or antiviral therapy available for the control of the disease, largely because of the lack of a permissive cell culture system for human noroviruses. The National institute of Allergy and Infectious Diseases classified noroviruses as category B pathogens because of their highly contagious

nature and potential to cause serious public health issue. Considering the above discussed challenges, antiviral drugs for controlling norovirus infections are in urgent need.

In vitro translation and mutagenesis studies indicated that the 200 kDa polyprotein of noroviruses is cleaved by the action of a 3C-like protease to produce three separate functional protein products (66). The active amino acid residues in the catalytic site of Norwalk virus protease are Cys139, His30, and Glu54. Cys 139 acts as nucleophile **(Figure 5.1)**. The 3C-like protease of Norwalk virus protease cleaves Gln/Gly or Glu/Gly sequences. In this study Norwalk virus 3C-like protease is characterized by using a fluorogenic substrate with EDANS-DABCYL pair. For Norwalk virus two fluorogenic substrates one with 14-amino acids (edans-EPDFHLQGPEDLAK-dabcyl) substrate derived from P7-P71 and another substrate with 7-amino acids (edans-DFHLQGP-dabcyl) substrate derived from P5-P2' were tested. Both the substrates were cleaved by the protease, but the substrate with 7-amino acids was more efficiently cleaved indicating that the amino acids at P7, P6, P3', P4', P5', P6', and P7' in the substrate are not that much important in the process of substrate cleavage by Norwalk virus protease. The efficient cleavage of substrate P5-P2' over P7-P7' is supported by the chromogenic substrate results shown by Hussey et. al. Their experimental results showed that chromogenic substrate containing P1- P3 residues was not cleaved by the Southampton virus 3C-like protease, and the substrate containing P5- P1 residues was more efficiently cleaved compared to that of substrates with P6-P1 or P4-P1 (48). Hence the substrate with 7-amino acid residues is used for the enzymatic activity characterization of inhibitor studies of Norwalk virus 3C-like protease.

The Norwalk virus displayed increase in proteolytic activity with the increasing concentrations of protease. This explains the effect of concentration of enzyme during a reaction. Assay buffer conditions were optimized for Norwalk virus protease by testing different percentages of glycerol, and different concentrations of DTT, NaCl and at various

pH values. The optimized assay buffer conditions for Norwalk virus protease is 20 mM HEPES, 0 mM NaCl, 0 mM DTT, 0.4 mM EDTA, and 50% glycerol and pH 8.0.

Poliovirus is an enterovirus belongs to the family Picornaviridae. Other important members of this family that causes important human and animal diseases are hepatitis A virus (HAV), coxsackievirus (CV), human rhinovirus (HRV), and foot and mouth disease virus (FMDV). Poliovirus is the causative agent of poliomyelitis in humans. Poliovirus was identified as causative agent for poliomyelitis in 1908 by Landsteiner and Popper (88). Based on slight differences in the capsid protein, poliovirus is divided into three serotypes, PV1, PV2, and PV3. All three forms of poliovirus are highly infectious, but among them PV1 is the most infectious form in nature (93). This virus is transmitted by fecal-oral route. Most of the infections occur in the gastrointestinal tract and the virus is shed in the feces of the infected individual. Poliovirus is strictly a human pathogen (99), (84). In most cases of the infection only primary viremia is observed and less than 1% of infections only lead to paralytic poliomyelitis. Poliovirus polyprotein codes for two viral proteases $2A^{pro,}$ and $3C^{pro}$. The catalytic site of poliovirus 3C protease consists of Cys147, His40 and Glu71 as active amino acid residues. Cysteine147 acts as nucleophile in the process of substrate cleavage through proteolysis **(Figure 5.2)**. The 3C protease is the most specific of its class, and cleaves solely at Gln↓Gly amino acid pairs. The translated polyprotein is proteolytically cleaved at 9 of 12 specific processing sites by poliovirus 3C protease and is mainly responsible for the production of separate gene products. Poliovirus has two potential vaccines oral polio vaccine (OPV) and inactivated polio vaccine (IPV). Even though most of the polio eradication was achieved by using OPV and IPV there are some back drops associated with these vaccines (22), (130). Hence there is a need to develop antiviral drugs against poliovirus to control the outbreaks. The concept of antiviral therapy is not new for poliovirus infection (17). Here, in this study a 9-amino acid fluorogenic substrate (edans-EALFQGPLP-dabcyl) was used to

characterize the poliovirus 3C protease. The assay buffer conditions were optimized for poliovirus 3C protease with the help of fast and continuous FRET protease assay method. Generally poliovirus 3C protease cleaves the sessile bond between Gln-Gly pair, preference for P4 alanine and P2' proline residues (82). The substrate was efficiently cleaved by the poliovirus 3C protease. The optimized buffer conditions for poliovirus 3C protease were 20 mM HEPES, 0 mM NaCl, 0 mM DTT, 0.4 mM EDTA, 20% Glycerol, and pH 6.0.

Transmissible gastroenteritis virus (TGEV) is coronavirus that causes diarrhea in pigs, and piglets under the age of 3 weeks usually die from infection. The other members of the genus coronavirus are SARS-CoV, MHV, IBV, and HCoV-229E. Coronaviruses infects humans and multiple species of animals. No specific vaccines or drugs are available to control HCoV infection. Some vaccines remain several potential problems, even though they are developed against IBV, CCoV, and TGEV.

PRCV is a variant of TGEV with no enterophathogenic ability and acts like a naturally occuring vaccine against TGEV (146). TGEV viral polyprotein processing is mainly achieved by main protease or 3C-like protease of 33.1 kDa. The catalytic site (dyad) of transmissible gastroenteritis virus main protease consists of Cys144 and His41 as active amino acids **(Figure 5.3)**. Cys144 acts as nucleophile during the process of proteolysis. The replicase polyproteins are cleaved by the main protease at not less than 11 conserved sites involving Leu, Gln ↓ (Ser, Ala, Gly) sequences. To date, no protease inhibitor has been developed for TGEV other than existing PRCV attenuated vaccine. TGEV main protease was characterized by using 12-amino acid fluorogenic substrate (edans-VNSTLQSGLRKM-dabcyl). The substrate was efficiently cleaved by the protease between the P1-P1' scissile bond and the coronavirus main proteases display strong preferences for Gln at P1 position, Leu at P2 position (2). Similar fluorogenic substrate with 12-amino acids (edans-VNSTLQSGLRK (dabcyl)-M was cleaved by SARS-CoV main protease (19) leading to the assumption that the

cleavage sites are conserved in coronaviruses. This is supported by work done by Haitao et.al. (2005). They synthesized a fluorogenic substrate MCA-AVLQSGFR-Lys (Dnp)-NH2 based on the assumption that the substrate-binding sites are highly conserved among CoV main proteases. The substrate sequence was derived from P4-P5' of SARS-CoV main protease N-terminal auto processing site sequence and it has a sequence AVLQSGFRK. Among all the main proteases tested, IBV main protease displayed almost similar Km to that of SARS-CoV main protease and four other CoV proteases (TGEV, FIPV, MHV, and HCoV-229E) showed stronger binding affinity to the substrate than SARS-CoV main protease supporting the assumption that substrates of CoV main proteases are conserved (147). The optimized assay buffer conditions for transmissible gastroenteritis virus main protease were 20 mM HEPES, 0 mM NaCl, 0 mM DTT, 0.4 mM EDTA, 20% Glycerol, and pH 6.0.

Development of antiviral drugs represents another tool for outbreak control, reduced transmission used on their own and work as adjuvants to vaccines. Since viral proteases are essential for the processing of the precursor polyprotein to produce mature or functional viral proteins, the viral proteases are targeted for the development of antiviral agents. Hence, inhibition of the viral protease has the potency to prevent the viral replication; many protease inhibitors have shown promising therapeutic activities in preclinical trails in animals and early clinical trials in man for viral and parasitic infections (1). A large number of protease inhibitors on the market or in the process of development for viruses like HIV, HCV, and SARS-CoV are some of the examples (34), (56), and (133).

The results from the assay buffer optimization conclude that, even though all the above discussed proteases are cysteine proteases which share several common characteristics (chymotrypsin-like fold, a Cys residue as nucleophile in the active site); they showed some differences in their activity. For example, Norwalk virus 3C protease showed increased proteolytic activity with the increase in the percentage of glycerol, whereas poliovirus 3C

protease showed maximum activity up to 30 % of glycerol and further increase in glycerol percentage decreased the proteolytic activity. TGEV main protease behaved totally different that it showed stable proteolytic activity at all glycerol percentages.

After optimizing the assay buffer conditions, inhibitor specificities for proteases from Norwalk virus, poliovirus, and TGEV were done using commercially available standard protease inhibitors. The rapid screening of standard protease inhibitors (chymostatin, leupeptin, TPCK, TLCK, and antipain) disclosed that TPCK inhibited NV, PV, and TGEV proteases with varying strength, and chymostatin inhibited only NV protease.

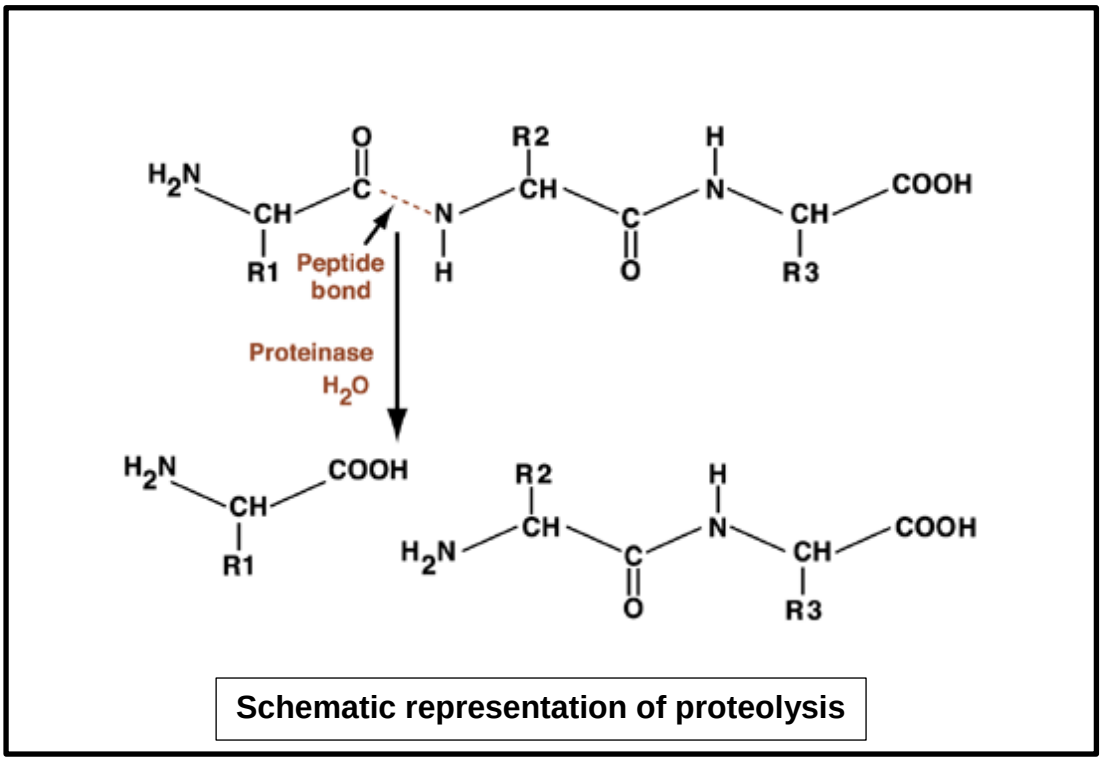
In summary, in this report, based on FRET assay, four fluorogenic substrates for detection of proteolytic activity of their corresponding proteases such as Norwalk virus protease, poliovirus 3C protease, and TGEV main protease were evaluated. Among the two substrates tested for Norwalk virus protease, the substrate with 7-amino acids showed much more efficiency than the substrate with 14-amino acids. Poliovirus 3C protease and transmissible gastroenteritis virus cleaved their substrates efficiently under optimized buffer conditions.

First, assay conditions of the FRET assay was optimized for each virus protease. Second, inhibition profiles of each virus protein were investigated using five commercially available standard protease inhibitors (chymostatin, leupeptin, antipain, TPCK, and TLCK). The inhibition studies showed that TPCK inhibited NV, poliovirus, and TGEV proteases with varying strength, and chymostatin inhibited only NV protease. All other inhibitors had little effects on the virus proteases. Finally, the present FRET based assay might supply an ideal dais for the large scale screening of inhibitors targeting Norovirus proteases, poliovirus 3C protease, and coronaviruses main proteases without time consuming.

FIGURES

Figure 1.1- The process of proteolysis:

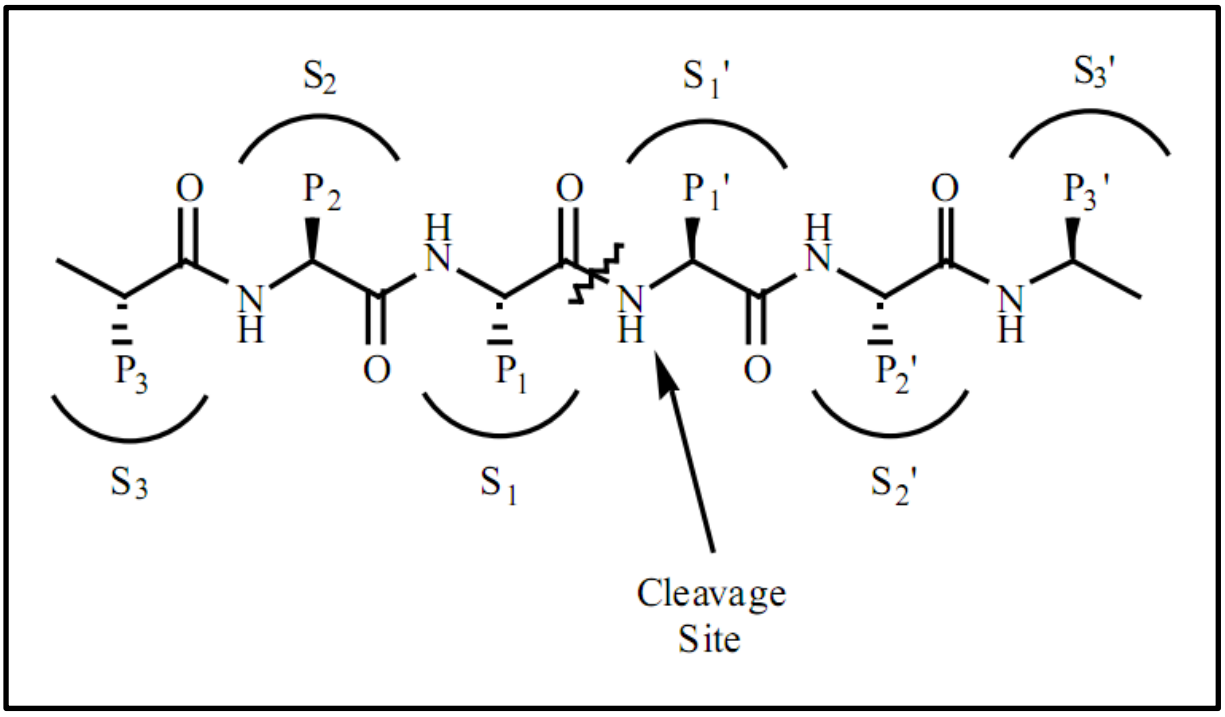
The figure shows the basic structure of a protein's two amino acids connected by a peptide bond represented with dotted line. The protease, by the addition of water cleaves the peptide bond and separates the protein into two fragments. The letters R1, R2, and R3 represent side chains of each amino acid.



http://pubs.niaaa.nih.gov/publications/arh27-4/317.htm-Intracellular proteolytic systems.

Figure 1.2- Cartoon showing common substrate and substrate binding sites for proteases:

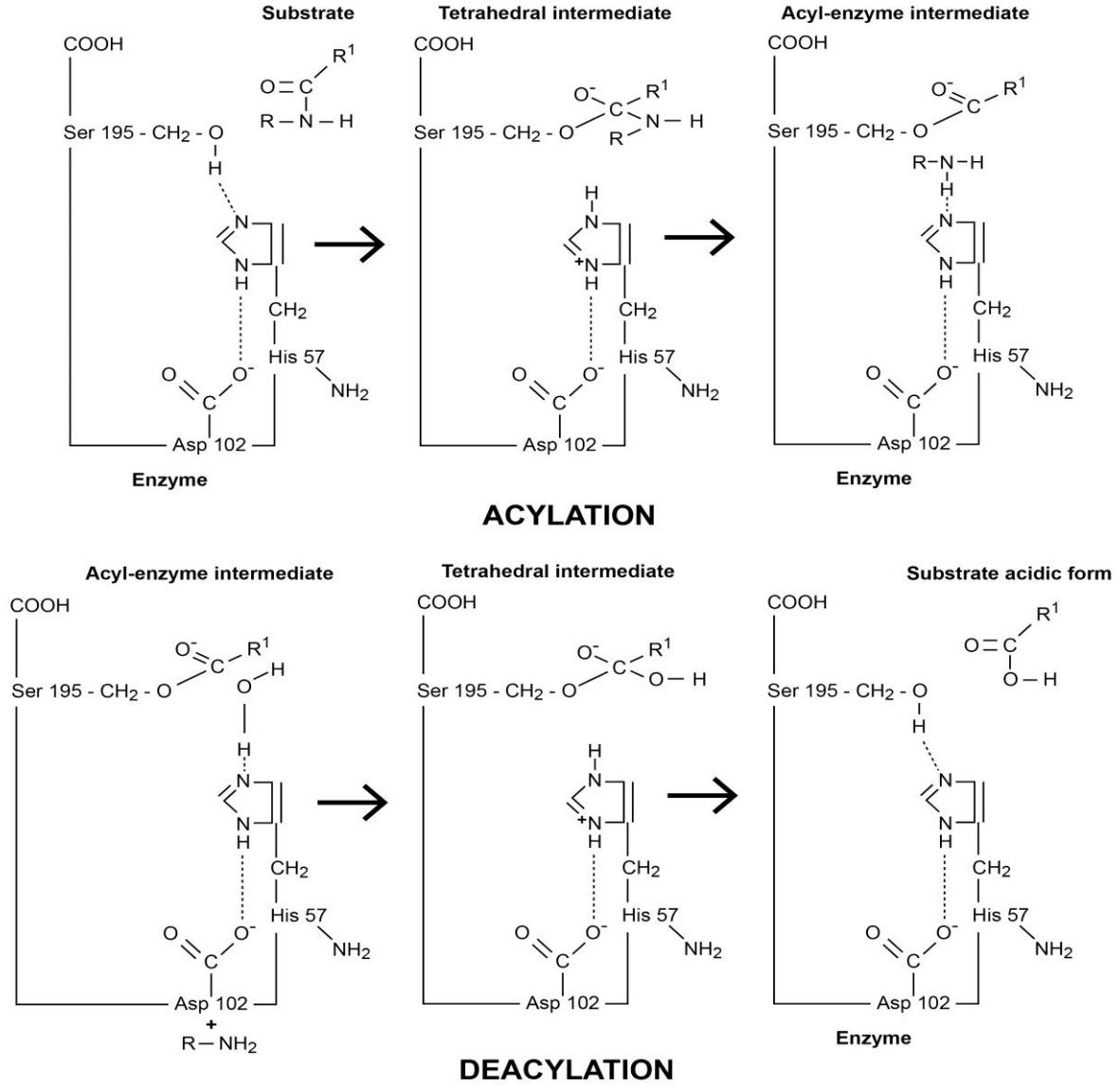
The cleavage site (scissile amide bond between P1 and P1') is indicated by an arrow. Amino acid residues towards the left from the cleavage site are called P1, P2, P3, and so on. In the similar manner, moving right from the cleavage site are residues P1', P2', P3', and so on. The corresponding substrate binding pockets on the protease for these residues are S1, S2, S3, and so forth on the left side and S1' S2', S3', and so forth on the right side.



Abbenante, G., and Fairlie, D. P., (2005). Protease inhibitors in the clinic. Medicinal Chemistry, 1, 71-104.

Figure 1.3- Serine protease catalytic mechanism:

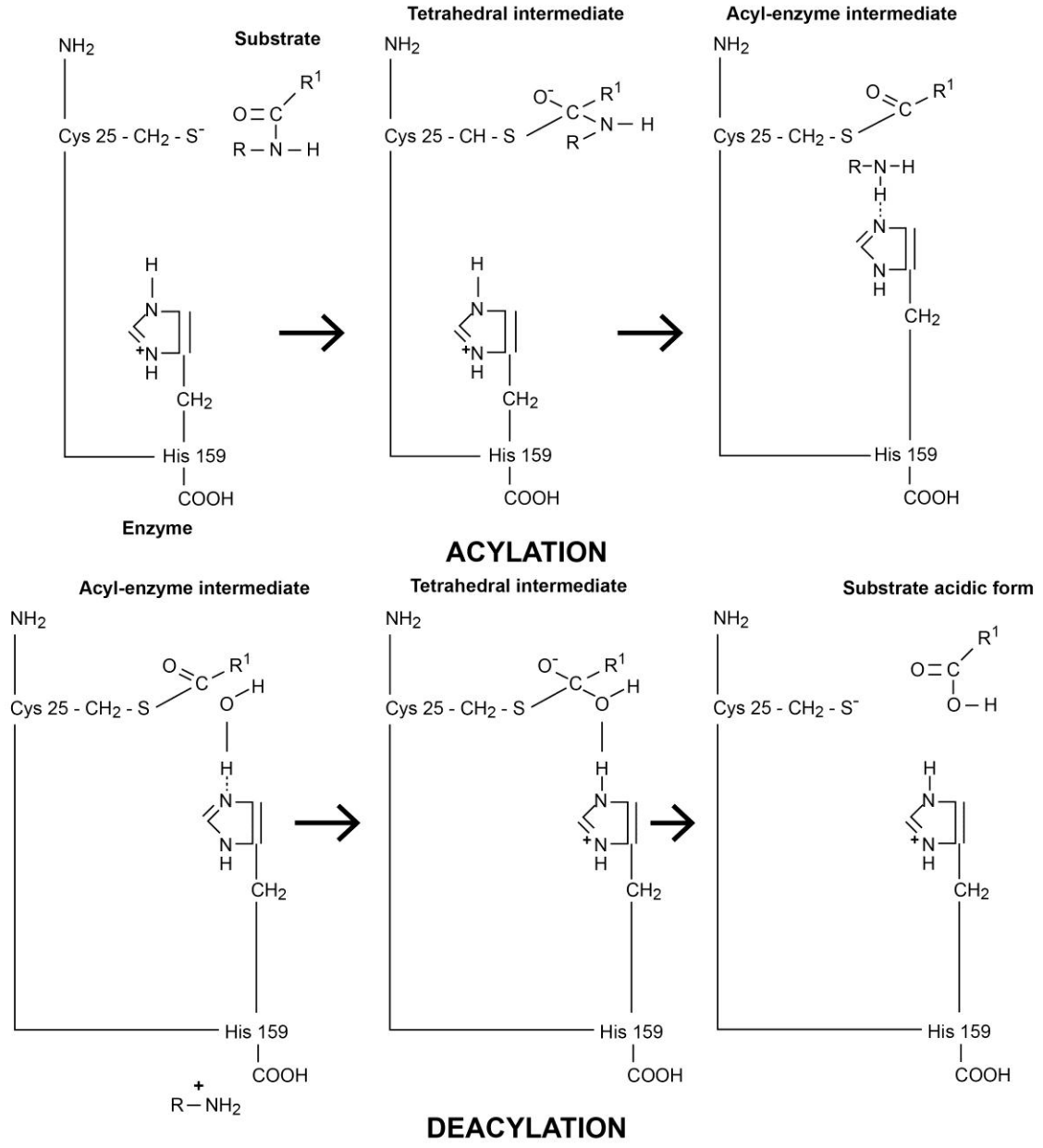
The serine protease catalytic mechanism is explained by using chymotrypsin as an example. The catalytic site consists of serine 195, histidine 57, and aspartic acid 102. In this mechanism, hydrolysis of substrate takes place in two stages. (i) In the first stage, a covalent acyl enzyme intermediate is formed by the nucleophilic attack of serine 195 of the enzyme (acylation). (ii) Secondly, the covalent acyl enzyme is breakdown by the attack of a water molecule releasing the carboxyl end of the peptide (deacylation). These two stages occur in the enzyme's active site with the active participation of histidine 57 and aspartic acid 102.



J Nduwimana et al.(1995). Proteases. Ann Biol Clin (Paris). 53(5):251-64.

Figure 1.4- Cysteine protease catalytic mechanism:

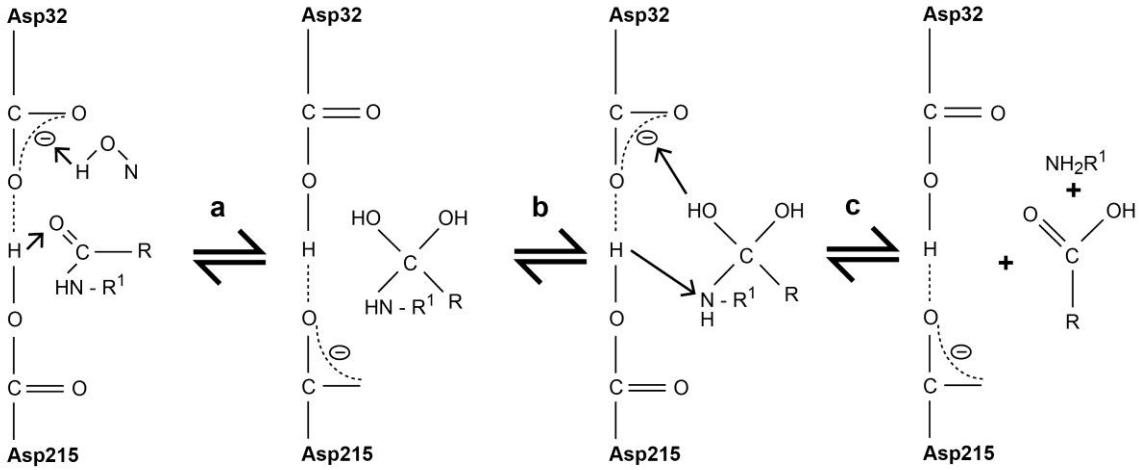
The cysteine protease catalytic mechanism is explained by using papain as an example. The catalytic site consists of cysteine 25, and histidine 159. In this mechanism, the hydrolysis of the substrate occurs in two stages: acylation and deacylation of the enzyme. (i) In the first stage, an acyl enzyme intermediate is formed by the nucleophilic attack of thiol (stabilized by imidazolium ion of the adjacent histidine 159) on the carbonyl group of the peptide bond to be cleaved. (ii) In the second stage, breakdown of acyl-enzyme intermediate by a water molecule releases the carboxyl end of the peptide.



J Nduwimana et al.(1995). Proteases. Ann Biol Clin (Paris). 53(5):251-64.

Figure 1.5- Aspartic acid proteases:

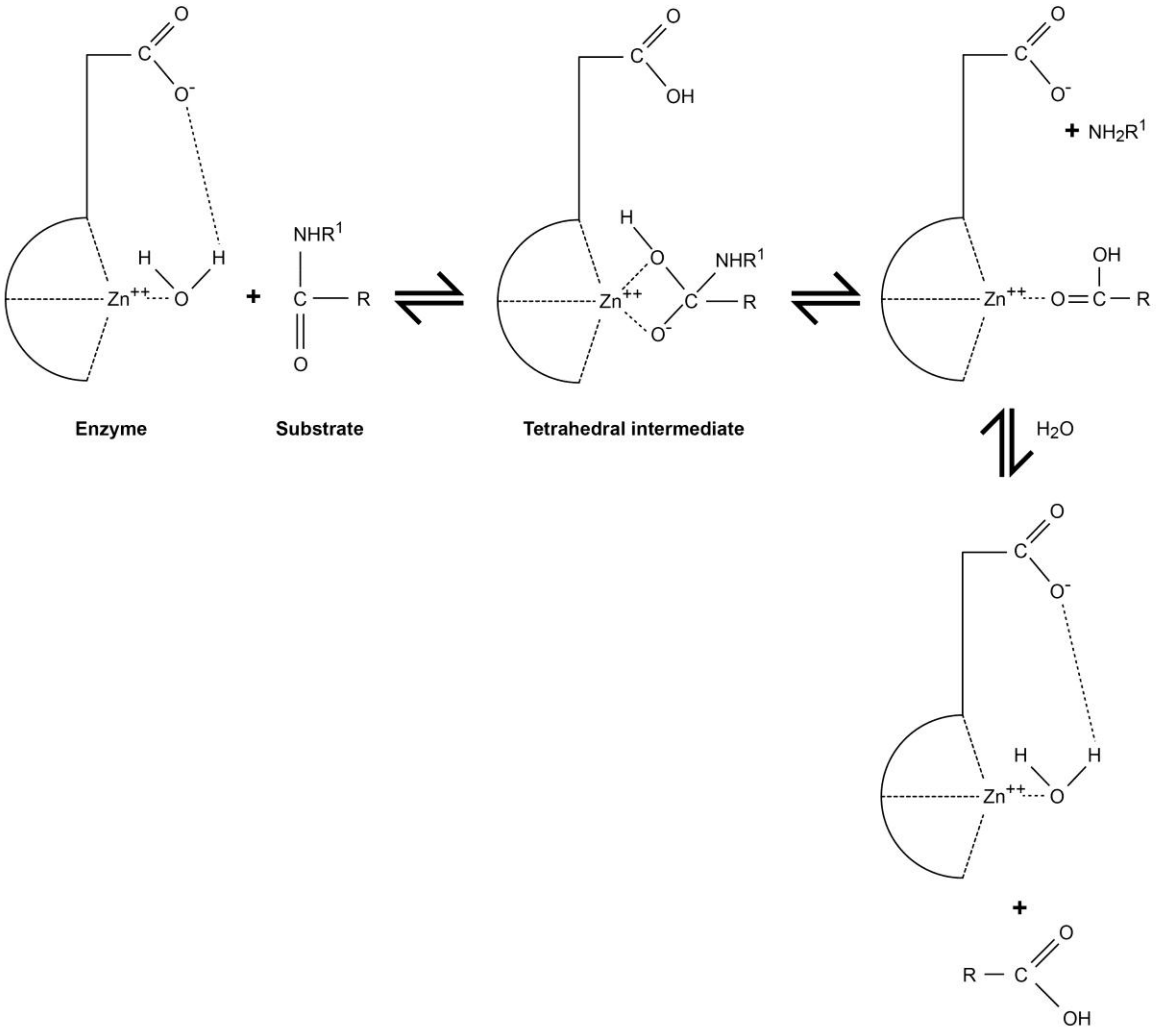
The aspartic acid protease catalytic mechanism is explained by using pepsin as an example. The catalytic site consists of two aspartic acid residues (aspartic acid 32 and aspartic acid 215) that are linked via a hydrogen bond. In case of aspartic acid proteases, two concurrent proton transfers lead to the formation of an impartial tetrahedral intermediate. First, a proton from a water molecule transfers to the carboxylate ion of aspartic acid 32, and second proton transfer occur from the carboxyl group aspartic acid 215 to the oxygen atom of the carbonyl group of the substrate. This tetrahedral intermediate is destroyed by the transfer of proton from intermediate to the carboxylate ion of aspartic acid 32, and transfer of another proton from carboxyl group of aspartic acid 215 to the nitrogen atom. This leads to the breakdown of C-N bond of the substrate.



J Nduwimana et al.(1995). Proteases. Ann Biol Clin (Paris). 53(5):251-64.

Figure 1.6- Metalloproteases catalytic mechanism:

The active site of metalloproteases consists of a metal ion, usually Zn^{2+} that is an integral part of the protease. Other than metal ion the active site always contains a Glu, sometimes associated with Trp or His. The carboxylate ion of the glutamic acid in the active site helps the zinc bound to water to make nucleophilic attack on the carbonyl carbon of the substrate to form a tetrahedral intermediate. Now carboxylate ion of the glutamic acid transfers the proton to the eliminated NH_2R' group which breaks down the tetrahedral intermediate finally releasing the carboxylic acid with the help of a water molecule.



J Nduwimana et al.(1995). Proteases. Ann Biol Clin (Paris). 53(5):251-64.

Figure 2.1- Diagrammatic representation of norovirus genome organization:

Norovirus genome consists of a molecule of single stranded positive sense RNA of approximately 7700 nucleotides with a polyadenylated 3' terminus. The genome is organized into three open reading frames (ORFs): ORF1 located towards the 5' end encodes a large 200 kDa non-structural precursor polyprotein, ORF2 and ORF3 located towards the 3' end encodes the major (VP1) and minor (VP2) proteins respectively.

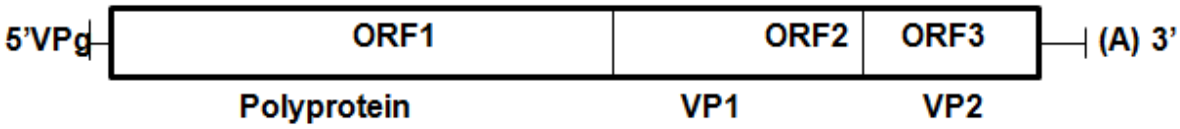


Figure 2.2- Proteolytic cleavage map of the non-structural proteins of Norwalk virus:

The 200kDa ORF1 polyprotein is cleaved by the action of 3C-like protease to initially generate three separate functional protein products. Full processing of ORF1 precursor polyprotein yields six mature non-structural proteins: (i) N-terminal protein (p45), (ii) an NTPase (p40), (iii) a 3A-like protein (p22), (iv) a VPg protein (p16), (v) a 3C protease, and (vi) an RNA-dependent RNA polymerase (p57). Specifically the protease cleaves at a Gln↓Gly or Glu↓Gly pair in the 200 kDa polyprotein ("↓" indicates the cleavage site). The protease preferential cleavage sites and the functions of the mature viral proteins are indicated in the diagram.

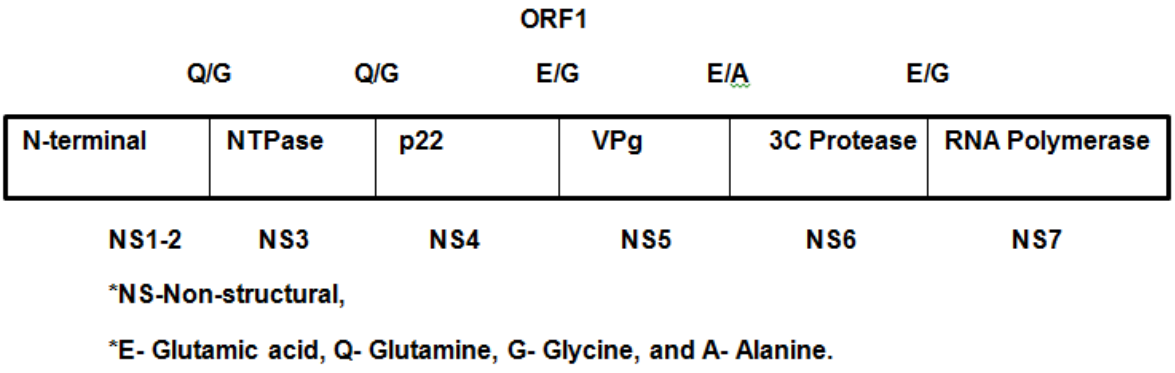
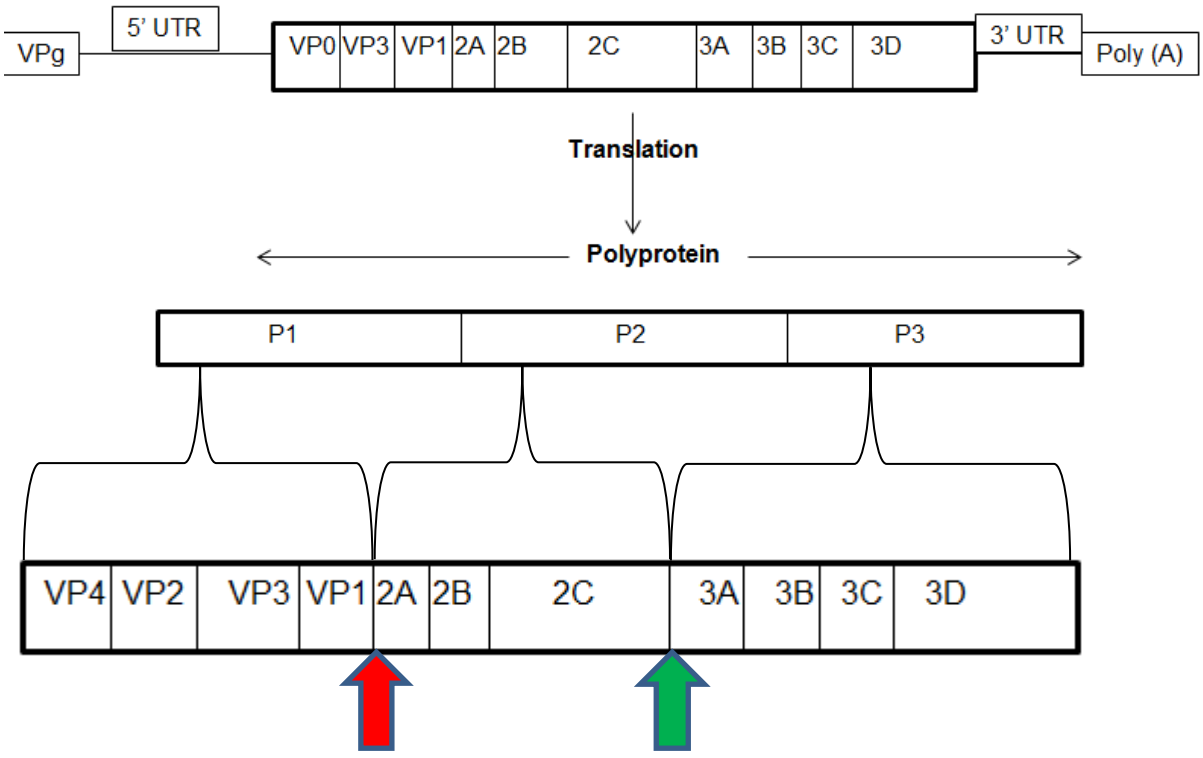


Figure 2.3 - Diagrammatic representation of poliovirus genome organization and primary cleavage sites of poliovirus polyprotein:

The genome of poliovirus is approximately 7400 nucleotides in length and composed of a single stranded positive sense RNA. It has a VPg at the 5' end linked to 5' untranslated region (UTR), a single open reading frame that encodes a polyprotein and a 3' untranslated region followed by a poly (A) tail. The polyprotein is divided into three regions, P1, P2, and P3 by two virally encoded proteases ($2A^{pro}$ and $3C^{pro}$). The preferential cleavage site for 3C protease is Gln/Gly pair, whereas 2A protease cleaves at a Tyr/Gly pair. 2A protease cleaves P1 and P2 (indicated by red arrow) and on the other hand 3C protease cleaves P2 and P3 (indicated by green arrow). The translated polyprotein is cleaved by 3C protease at 9 of 12 specific processing sites.



Mature functional proteins.

Figure 2.4 – Genomic organization of transmissible gastroenteritis virus:

The genome of TGEV is a single stranded positive sense RNA of approximately 28, 500 nucleotides in length and consists of seven open reading frames. ORF1 and ORF2 together constitute gene 1, which comprises about two thirds of the genome from the 5' end. It has a leader sequence at the 5'end followed by untranslated region. Another UTR is present at the 3'end followed by poly (A) tail.ORF1a and ORF1b encode polyprotein 1a and polyprotein 1b respectively by ribosomal frame shifting. The processing of these polyproteins produces functional proteins that are required for the viral transcription and translation. ORF1 encodes two papain like proteases (process the N-terminal region) and a chymotrypsin like protease (Main protease) which is responsible for the processing of remaining polyprotein. The remaining six ORFs code for structural proteins spike (S), envelope (E), transmembrane (M), Nucleocapsid (N).

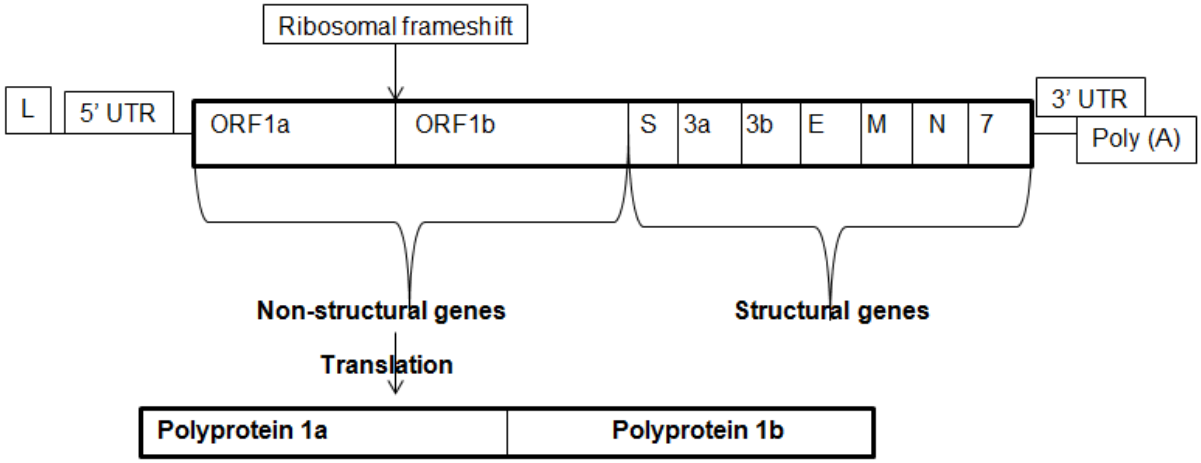


Figure 2.5- Fluorescence resonance energy transfer (FRET) assay:

In FRET based assays, a substrate is designed on the basis of a sequence that is recognised and cleaved by the protease. FRET substrates have a fluorescent donor on one side and a fluorescent quencher on the other side of the cleavage site. The donor fluorescence signal is quenched by the nearby fluorescent quencher in an uncleaved substrate. In the presence of a protease, if the substrate is cleaved, the donor will get separated by the quencher and thereby increasing the fluorescence yield of the donor group. Addition of inhibitors will decrease the intensity of donor fluorescence signal

.

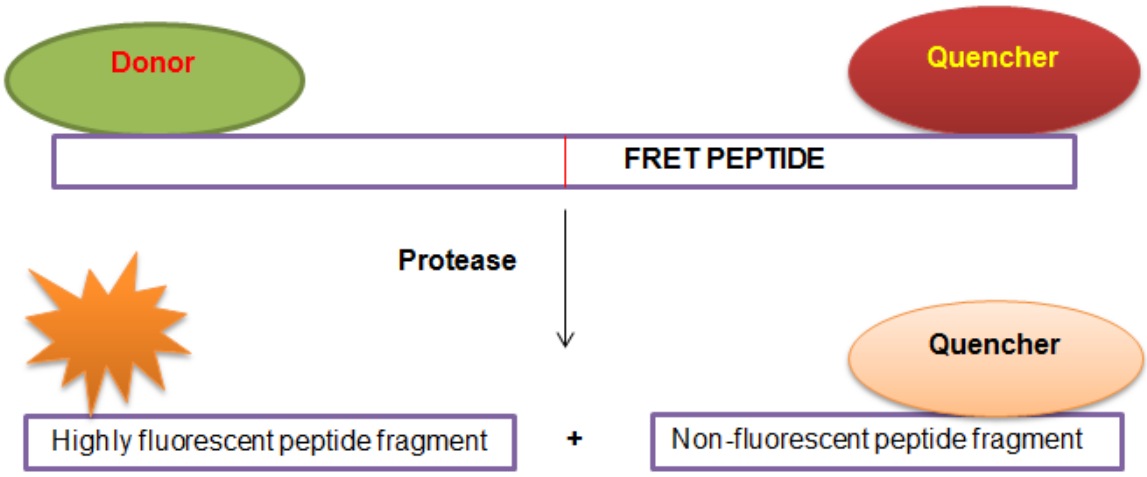


Figure 4.1- SDS-PAGE analysis of purified proteases:

The SDS-PAGE analysis results of purified proteases from Norwalk virus, poliovirus, and transmissible gastroenteritis virus. The Norwalk virus protease has a predicted molecular weight of 19 kDa, poliovirus protease has a predicted molecular weight of 20 kDa, and TGEV protease has a predicted molecular weight of 33 kDa. In the figure, lane 1 corresponds to molecular weight EZ-marker, lane 2- Norwalk virus protease, lane 3- poliovirus protease, and lane 4- TGEV protease.

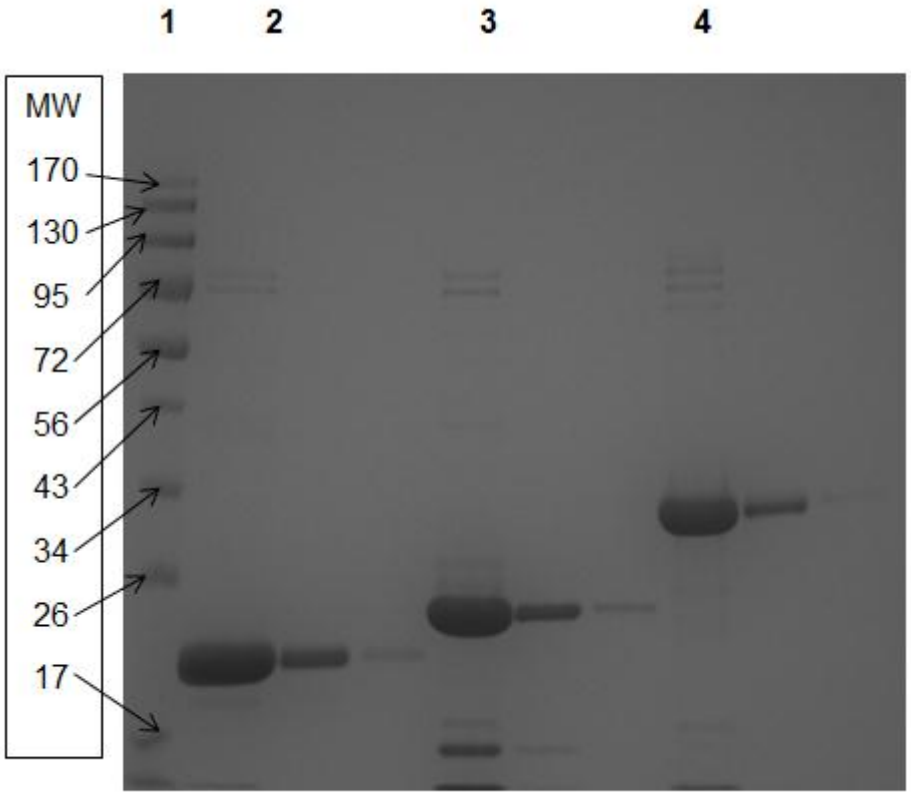


Figure 4.2- Optimization of protease concentrations for Norwalk virus, poliovirus and TGEV:

Different concentrations of proteases are tested with same concentrations of their respective substrates. The concentration of substrate used is 20 µM per reaction. A protease concentration for each protease is showed on their respective graph. For every protease, increase in the concentration of protease increased the enzymatic activity. Here, in this figure, **(A)** corresponds to Norwalk virus protease, **(B)** corresponds to poliovirus protease, and **(C)** corresponds to TGEV protease.

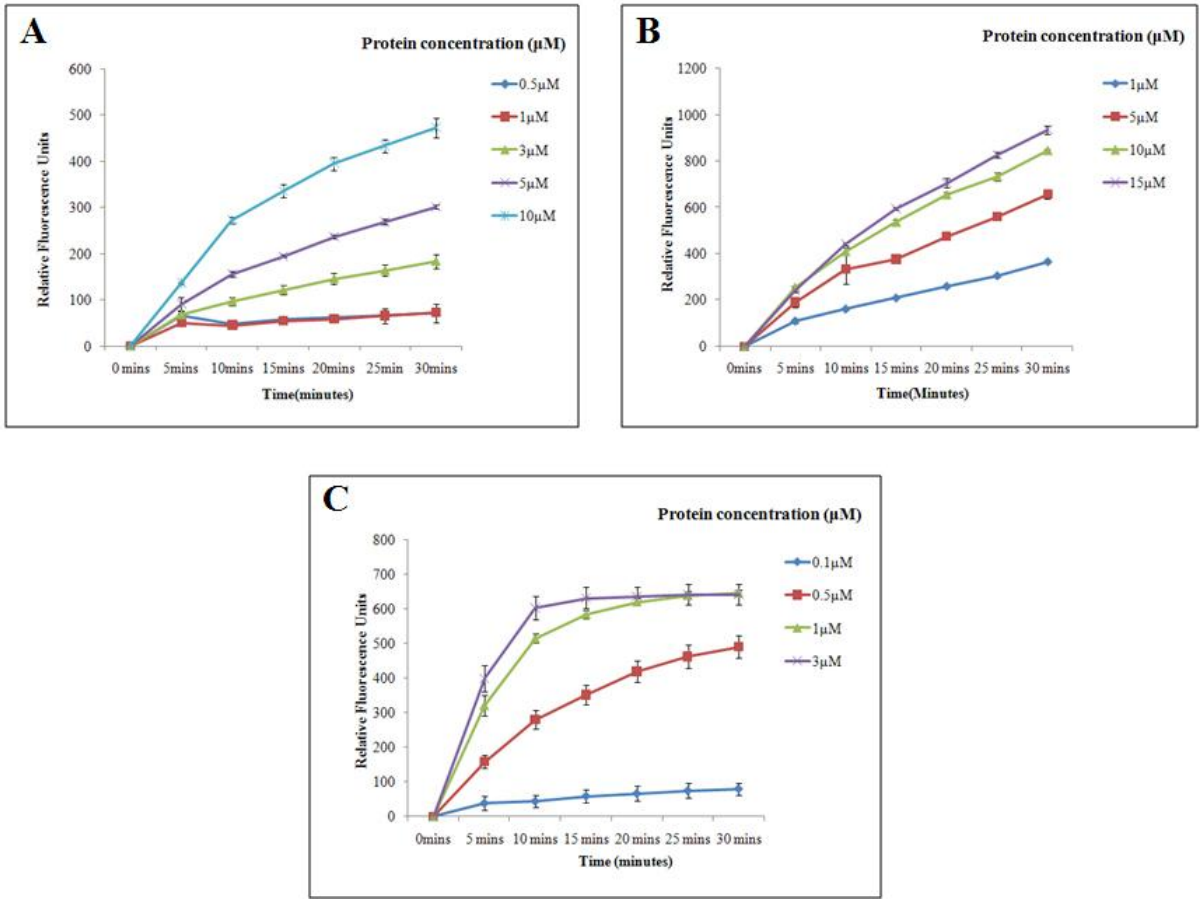


Figure 4.3- Optimization of glycerol percentage in assay buffer:

The percentage of glycerol is optimized for every protease in this study. A range of glycerol percentage from 20 % to 50% was tested for each protease. Norwalk virus protease remarkably showed increased enzymatic activity with the increase in the percentage of glycerol. Poliovirus showed maximum activity at 20 % and 30%. Further increase in glycerol percentage decreased the activity of poliovirus protease. Interestingly TGEV virus showed stable activity at all percentages of glycerol tested. In this figure, **(A)** is Norwalk virus protease activity at different glycerol percentages, **(B)** is poliovirus protease, and **(C)** is TGEV protease activity at different glycerol percentages.

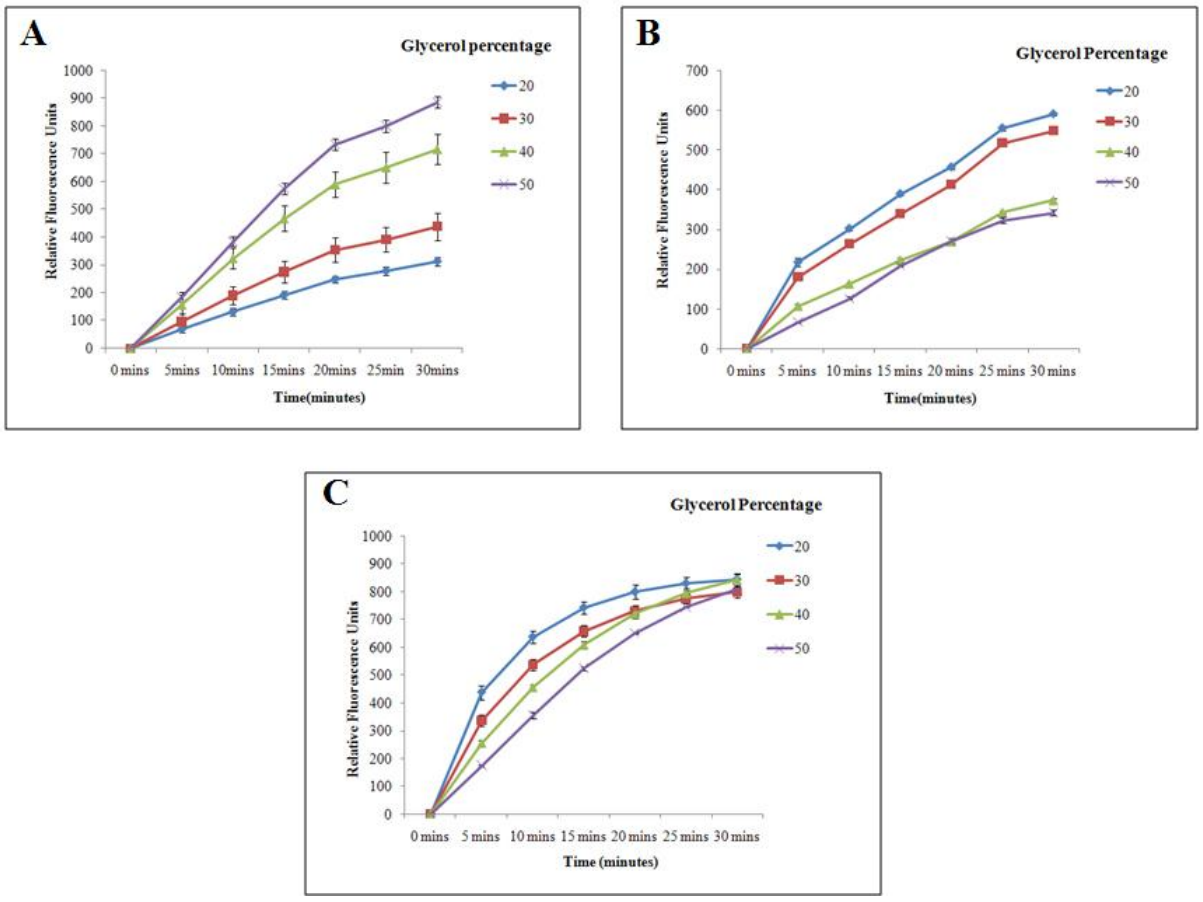


Figure 4.4- Optimization of DTT concentration in assay buffer:

Different concentrations of DTT from 0 mM to 30 mM in the assay buffer are tested for each protease. There is no change observed between 0 mM to 10 mM DTT concentration. Norwalk virus and poliovirus proteases showed reduced activity at 30 mM concentration of DTT, whereas TGEV protease showed stable activity at 30 mM concentration of DTT also. In this figure DTT concentration values corresponding to 0 mM and 30 mM are displayed because there is no differential change observed between 0 mM and 10 mM concentrations of DTT. In this figure, **(A)** represents Norwalk virus protease, **(B)** represents poliovirus protease, and **(C)** represents TGEV protease.

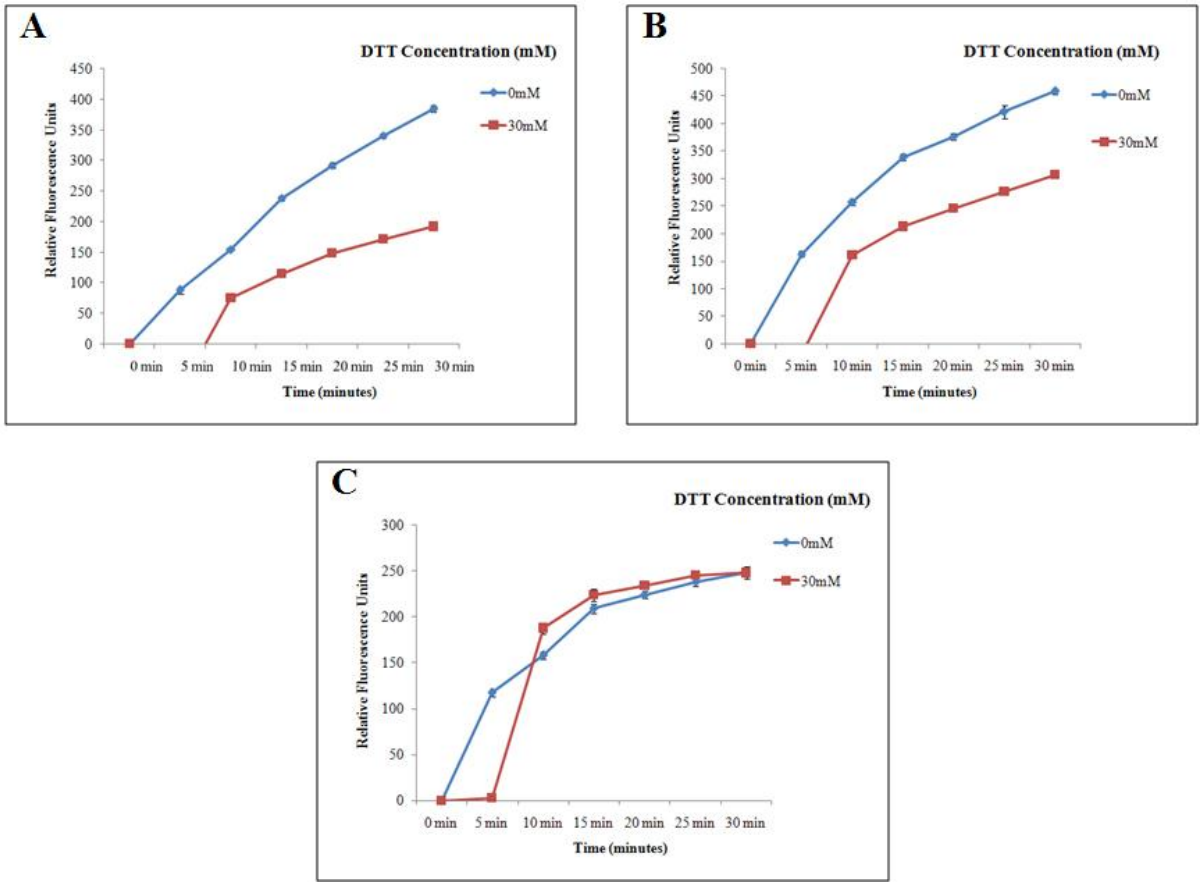


Figure 4.5- Optimization of NaCl concentration in the assay buffer:

A different concentration of NaCl (0, 50, 100, 200, and 400 mM) in assay buffer was tested for each protease. **(A)** Norwalk virus showed maximum enzymatic activity at low concentrations of NaCl, even better without NaCl in the buffer. **(B)**Poliovirus demonstrated almost similar kind of activity at all concentrations of NaCl tested. **(C)** TGEV showed maximum proteolytic activity without NaCl in the assay buffer and increase in the concentration of NaCl decreased the activity of TGEV protease.

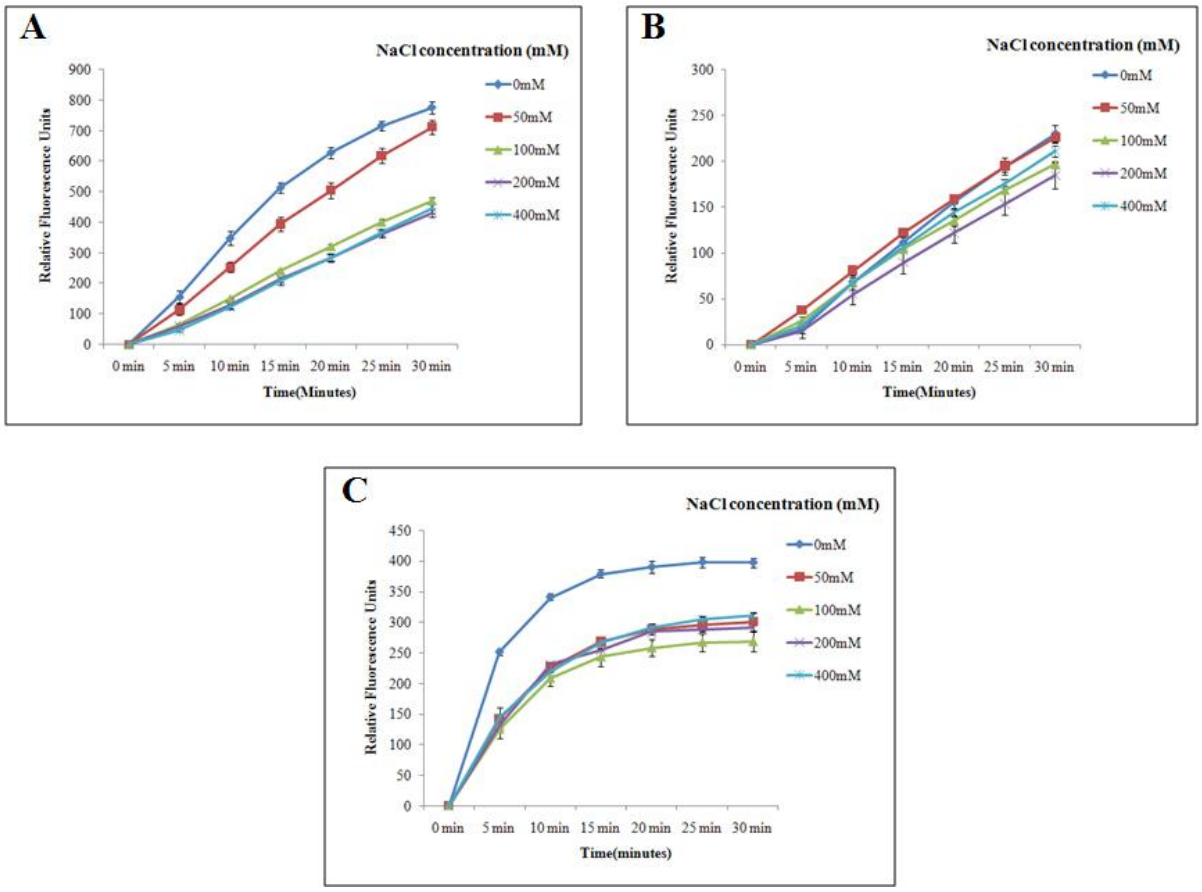


Figure 4.6- pH range optimization:

A range of pH values from 6 to 10 are tested for each protease. **(A).** Norwalk virus protease showed maximum enzymatic activity at pH 7 to 9. **(B).** Poliovirus protease showed maximum activity from pH value 6 to 8, but showed more activity at pH 6. **(C)**. TGEV protease preferred pH 6 for its maximum enzymatic activity and at pH 7, its activity is reduced by 50% showing that further increase in pH will decrease the enzymatic activity of TGEV.

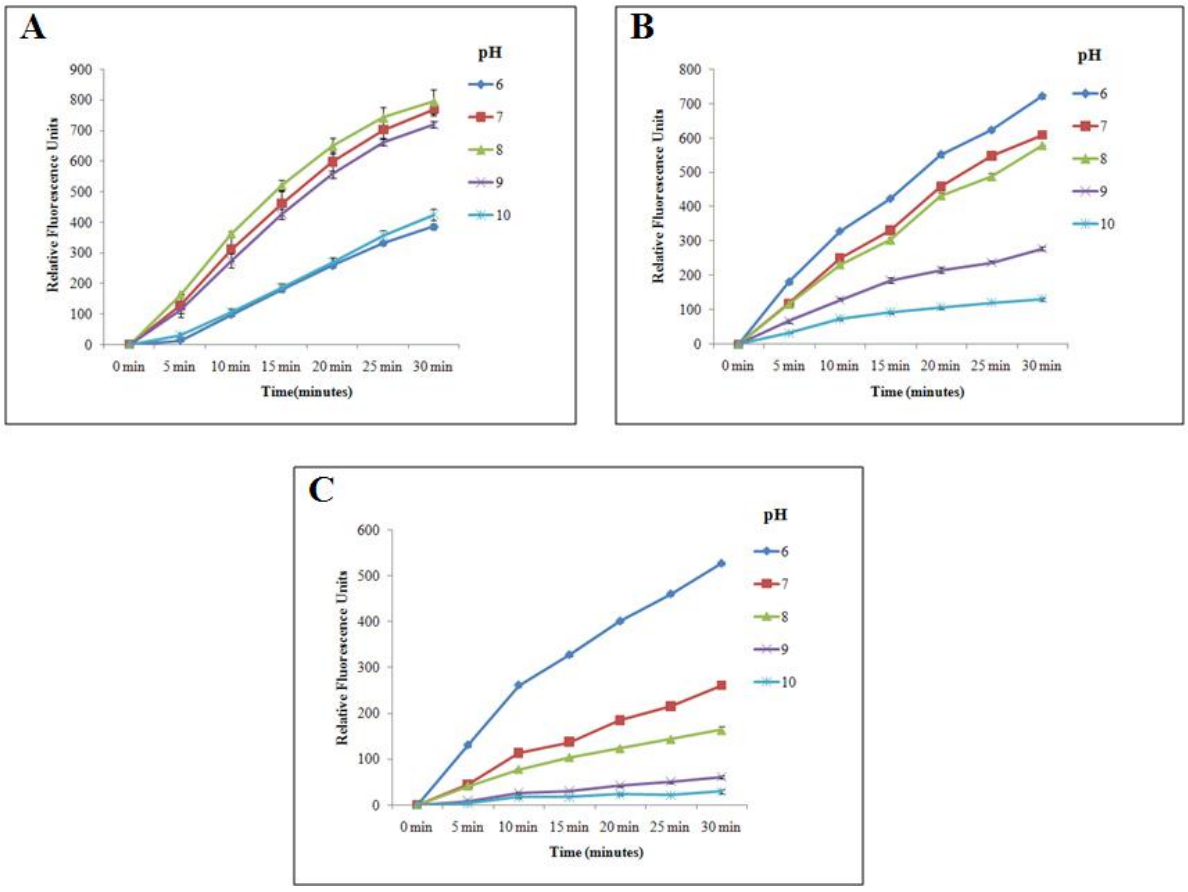


Figure 4.7- Rapid screening of protease inhibitors:

All the three proteases were tested against five commercially available protease inhibitors namely chymostatin, leupeptin, antipain, TPCK, and TLCK. The concentrations of inhibitors tested were 50 µM and 100 µM respectively for each protease. **(A).** Norwalk virus was maximum inhibited by chymostatin at 50 µM and 100 µM concentrations with very slight differences. **(B).**Among the five inhibitors tested against poliovirus protease, TPCK at 100 µM concentration inhibited the poliovirus protease more efficiently than 50 µM of TPCK. **(C).** TGEV protease was partially (i.e. less than 50%) inhibited by all the protease inhibitors tested and failed to show strong inhibition with any protease inhibitor.

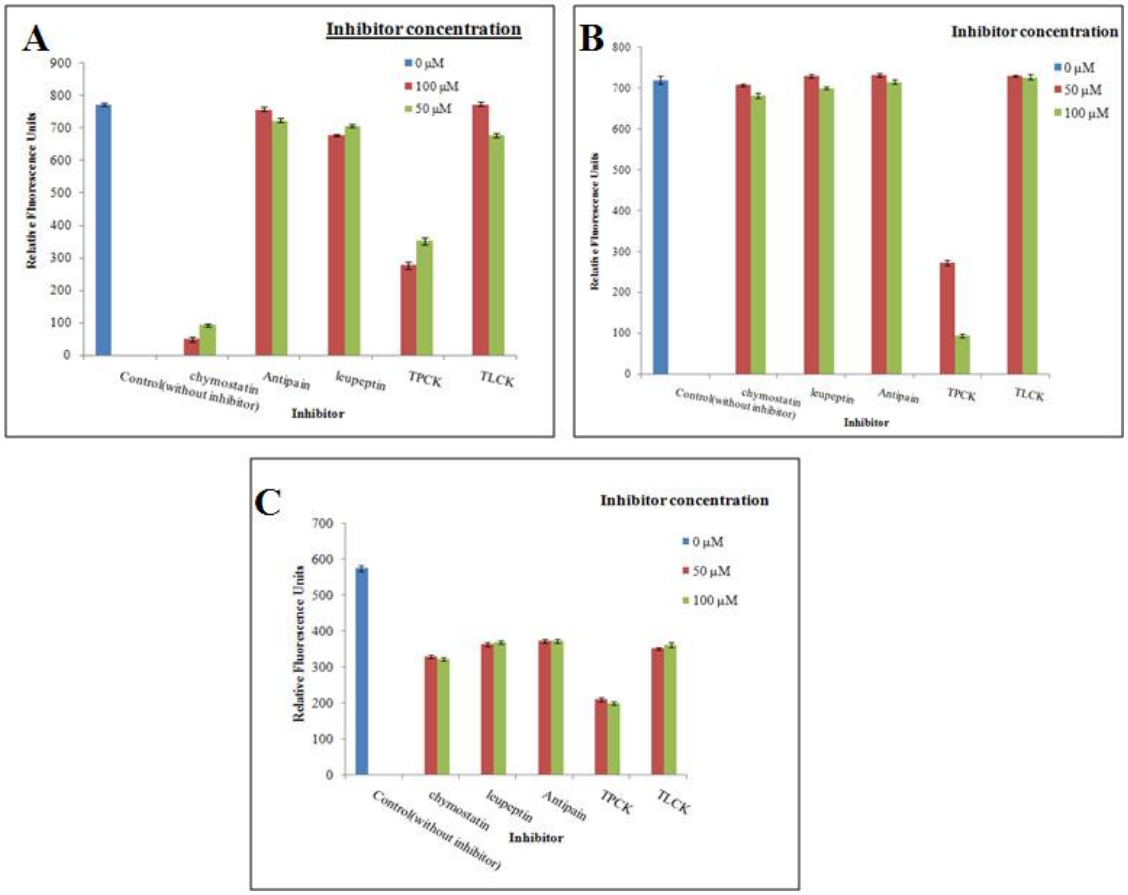


Figure 4.8- Dose response curves of chymostatin on Norwalk virus protease and TPCK on poliovirus protease:

(A). Chymostatin at 50 μM concentration is used for getting dose response curve because chymostatin at 50 μM concentration showed strong inhibition with Norwalk virus protease. Hence, different concentrations of chymostatin (0, 10, 20, 30, 40, and 50 μM) are used to calculate IC_{50} value for chymostatin. **(B).** similarly for poliovirus protease, TPCK at 150 μM concentration is serially diluted to calculate the IC_{50} value.

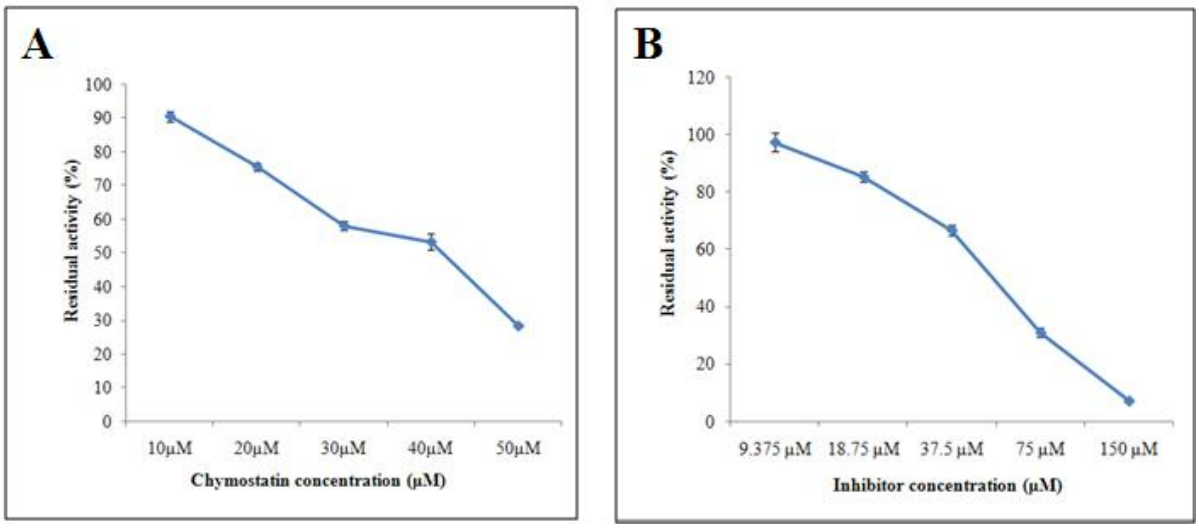


Figure 5.1A- The crystal 3D structure of Norwalk virus protease:

The crystal structure of Norwalk virus protease is shown in the figure. The N-(fibre brick) and C-terminal (yellow) domains are shown in the figure. Two domains are connected by a loop (lime green). β – Strands and helices are represented as arrows and ribbons. The three active site amino acid residues [Cys139 (cyan), His30 (blue), and Glu54 (green) are labeled and represented in the figure. (PDB ID-2FYQ)

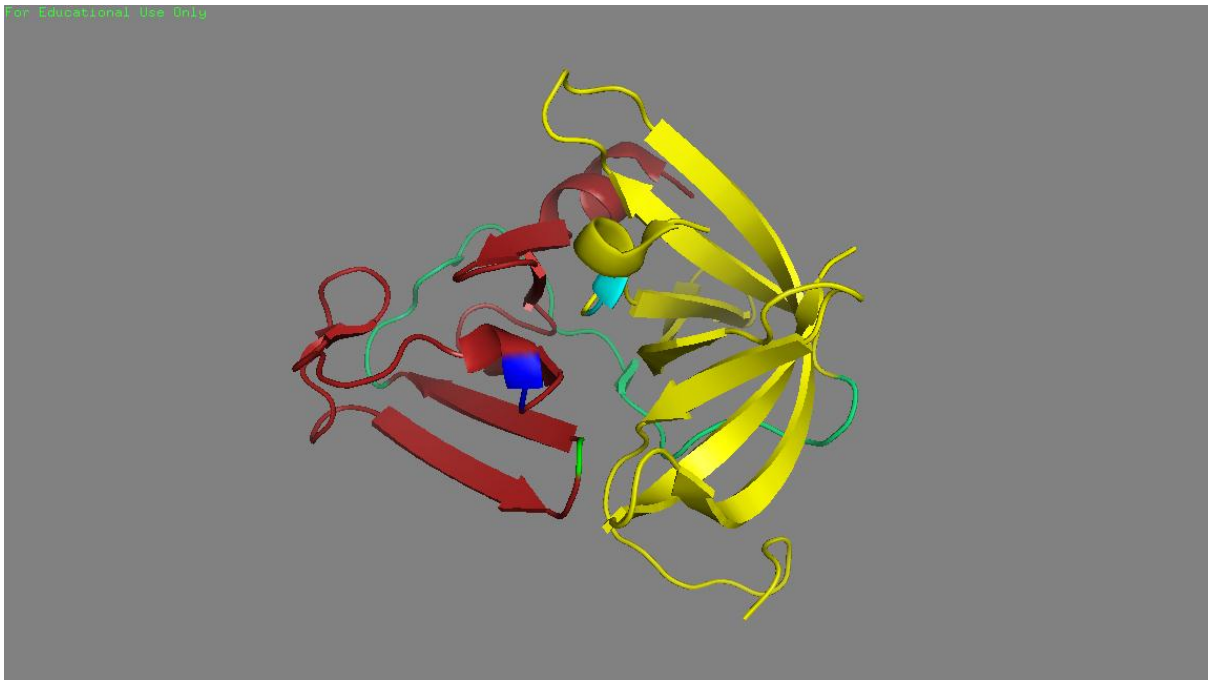
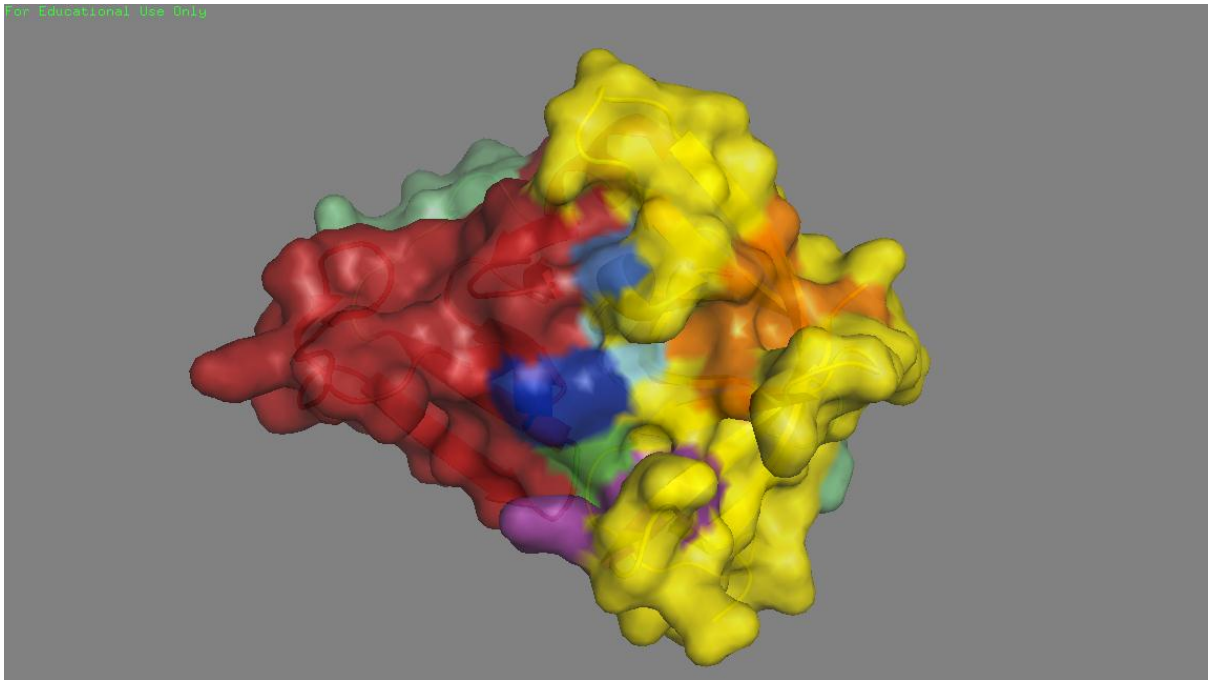


Figure 5.1B- Substrate binding sites on the surface of Norwalk virus 3C protease:

The surface representation of Norwalk virus 3C protease is shown in the figure with 20% surface transparency. In the figure, the S1 specificity pocket is shown in orange color; S2 specificity pocket is shown in purple color and the oxyanion hole is shown in marine color. The N-terminal domain is represented by fibre brick color and the C-terminal is represented by yellow color. The active amino acids Cys139 is shown in cyan; His30 is shown in blue, and Glu54 is shown in green color respectively (PDB ID-2FYQ).



The PyMOL molecular graphics

Figure 5.1C- The crystal structure of Southampton virus 3C protease linked to an active site directed peptide inhibitor:

The crystal structure of Southampton virus 3C protease is shown in the figure. The N-(fibre brick) and C-terminal (yellow) domains are shown in the figure. Two domains are connected by a loop (lime green). β – Strands and helices are represented as arrows and ribbons. The three active site amino acid residues [Cys139 (cyan), His30 (blue), and Glu54 (green)] are labeled and represented in the figure. An active site directed peptide inhibitor (acetyl-Glu-Phe-Gln-Leu-Gln-X) which is based on the rapidly cleaved recognition sequence in the 200 kDa polyprotein substrate is also represented as sticks in red color.(PDB ID-2IPH)

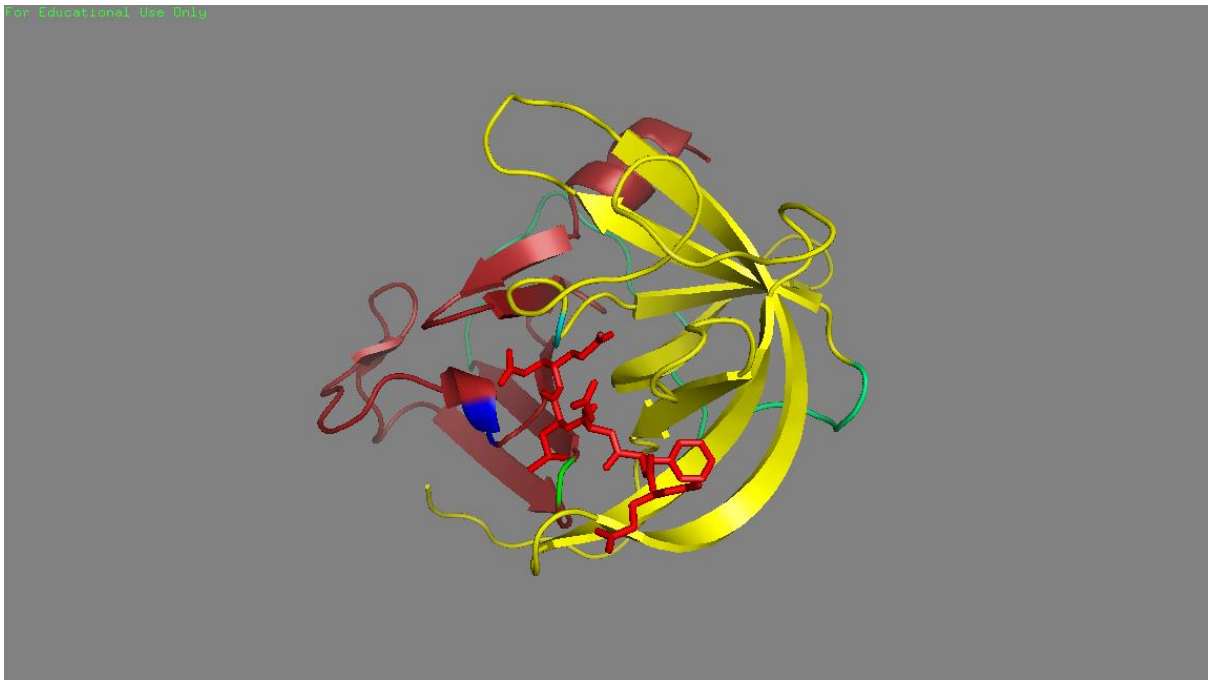


Figure 5.1D- Comparison of Norwalk Virus and Chiba virus 3C proteases:

The Norwalk and Chiba virus 3C proteases were aligned using PyMOL molecular graphics. The Norwalk virus is shown in cyan color and the Chiba virus is shown in warm pink color. Two proteases showed maximum alignment with each other. Active amino acid residues in both the proteases were represented sticks with same colours [Cys139 (green), His30 (orange), and Glu54 (yellow)].

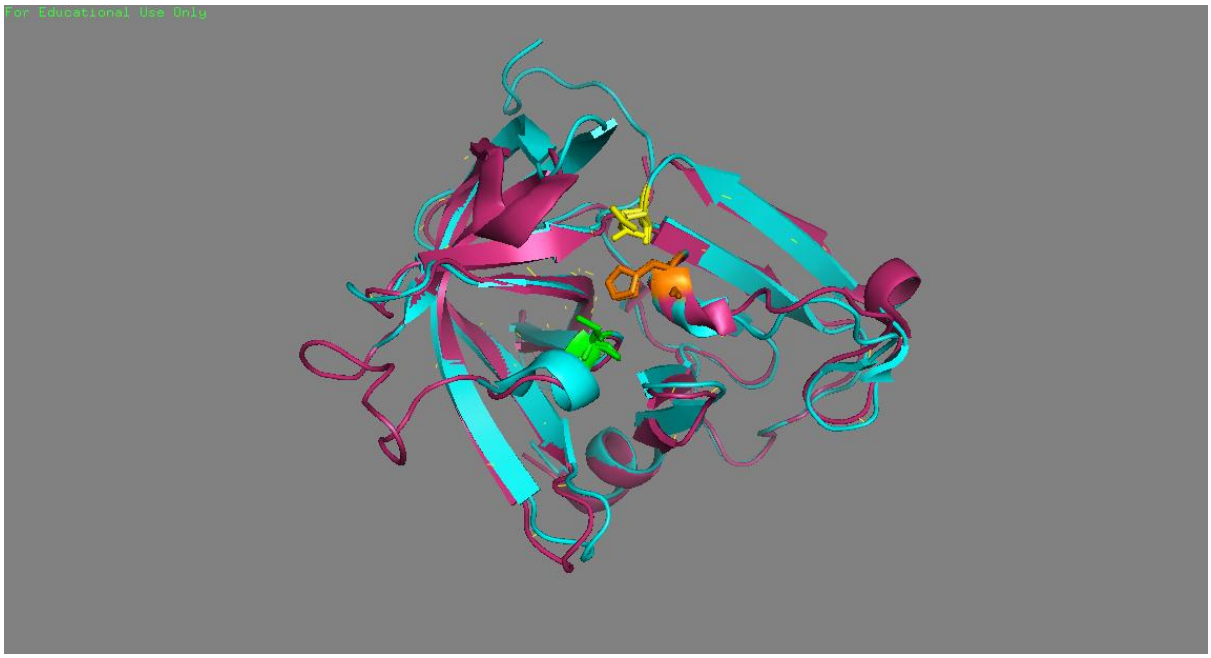


Figure 5.2A- The crystal 3D structure of poliovirus 3C protease:

The 3D diagram of poliovirus 3C protease showing the overall fold is represented here. It consists of two domains. Domain I is represented with warm pink and domain II is represented with yellow color respectively. β- Strands are indicated by arrows, helices are shown as ribbons. Both loop and turn structures are represented as coils. The poliovirus active site is comprised of a catalytic triad [Cys147 (green), His40 (cyan), and Glu71 (blue)], which is also shown in the diagram. The poliovirus 3C protease active site is similar to, but larger than the proteolytic sites of the chymotrypsin-like enzymes. (PDB ID-1L1N)

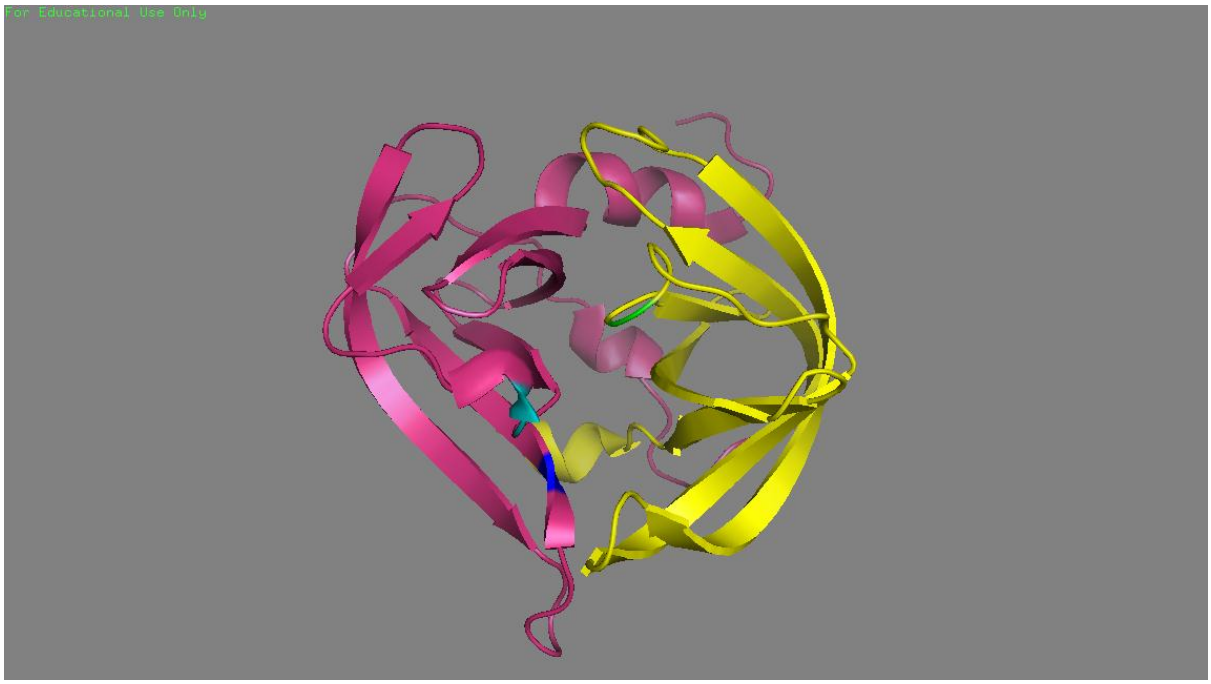


Figure 5.2B- Substrate binding sites on the surface of poliovirus 3C protease:

The surface representation of poliovirus 3C protease is shown in the figure with 20% surface transparency. In the figure, the S1 specificity pocket is shown in magenta color; S2 specificity pocket is shown in red color and the oxyanion hole is shown in orange color. The active amino acids Cys147 is shown in green; His40 is shown in cyan, and Glu71 is shown in blue color respectively (PDB ID-1L1N).

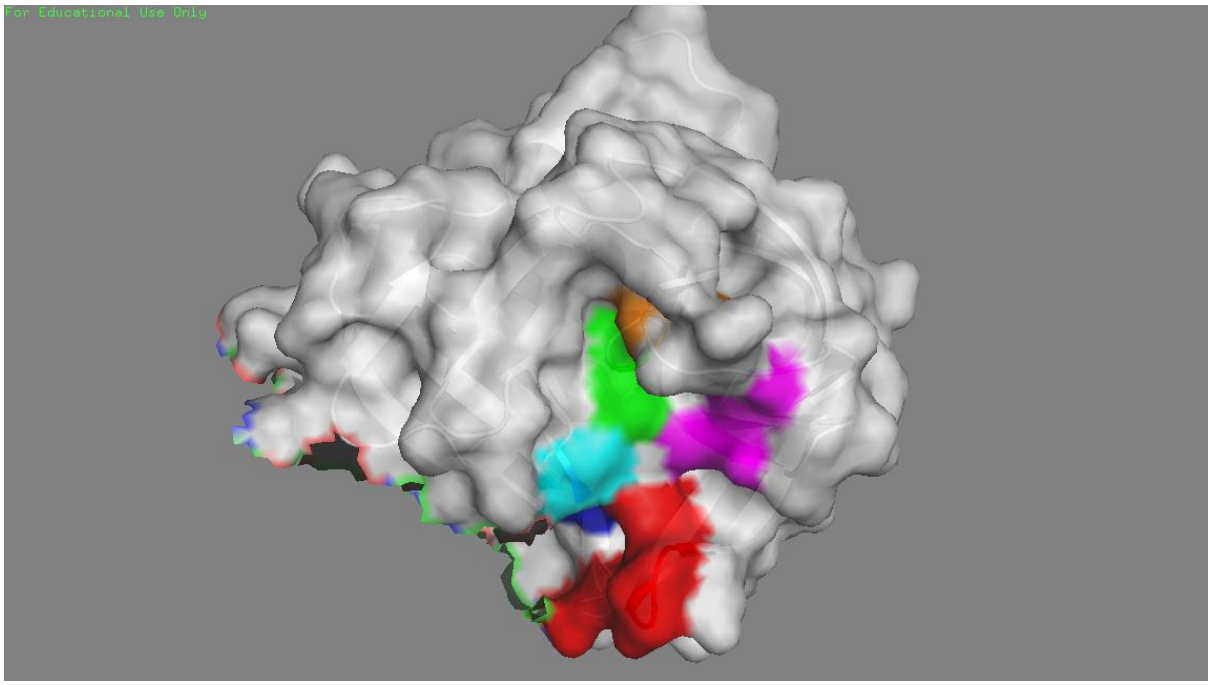


Figure 5.2C- Comparison of Norwalk Virus and poliovirus 3C proteases:

The Norwalk and poliovirus 3C proteases were aligned using PyMOL molecular graphics. The Norwalk virus is shown in cyan color and the poliovirus is shown in green color. The active site cleft of two proteases is conserved between each other, but the cleft of poliovirus 3C protease is slightly larger than Norwalk virus protease. Active amino acid residues in both the proteases were represented as sticks with similar colours. In case of Norwalk virus 3C protease Cys139 is shown in orange, His30 is shown in red and Glu54 is shown in magenta. Poliovirus Cys147 is represented with orange, His40 is represented with red and Glu71 is represented with magenta color.

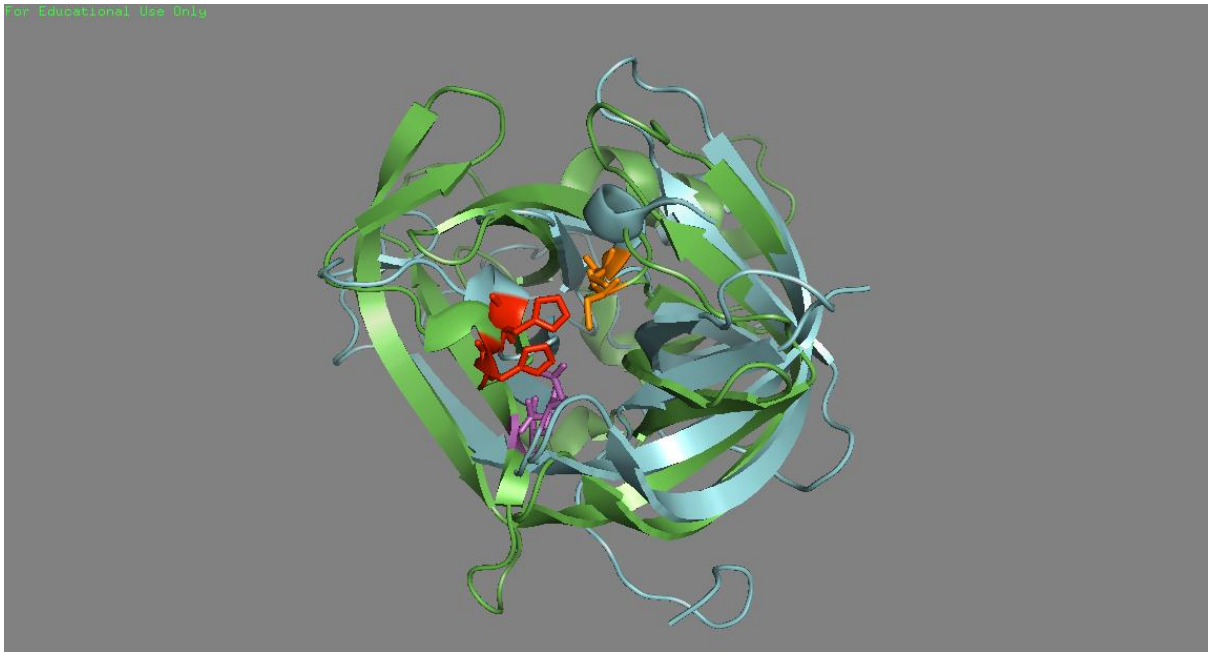


Figure 5.3A- The crystal 3D structure of transmissible gastroenteritis virus main protease:

The 3D structure showing the overall fold of TGEV main protease is shown here. TGEV main protease consists of two β – barrel domains and one α – helical C-terminal domain. The β – barrels of each domain I (yellow) and II (lime green) are composed of six-stranded β – sheets. Domain III (tv red) is mainly composed of α- helices. Domains II and III are connected by a long blue color loop. α –helices and β –strands are represented as coils and arrows, respectively. The active amino acid residues [Cys 144 (red) and His 41 (magenta)] are labeled and showed in the figure.

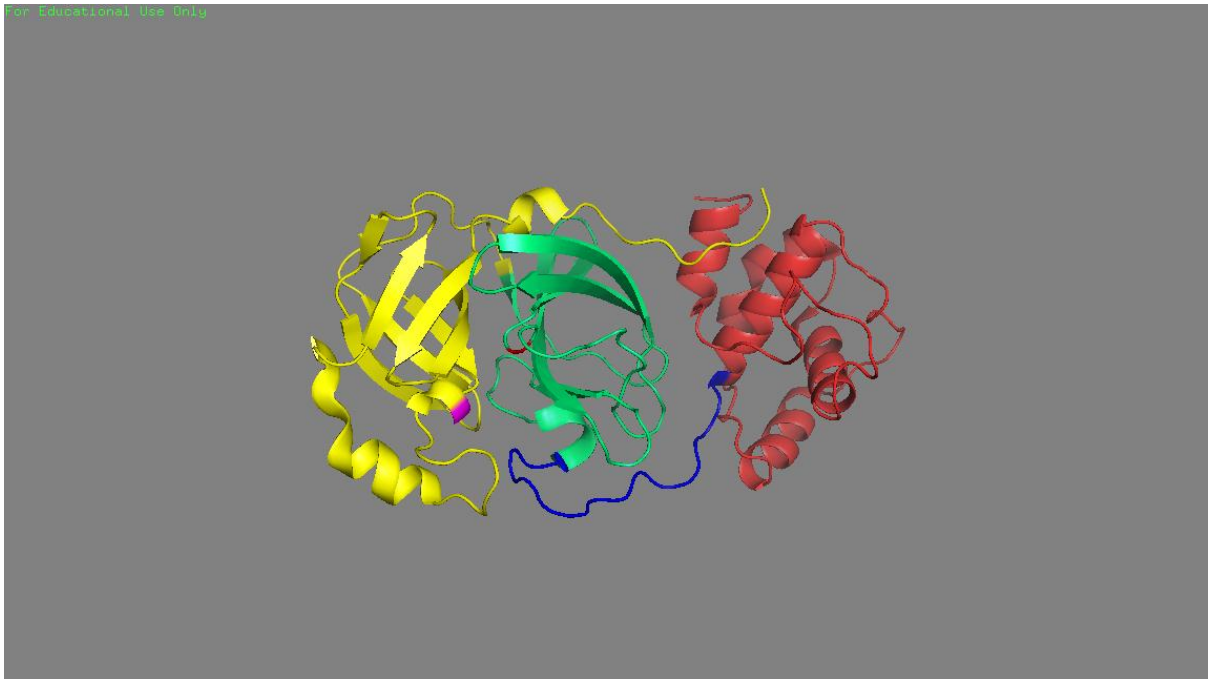
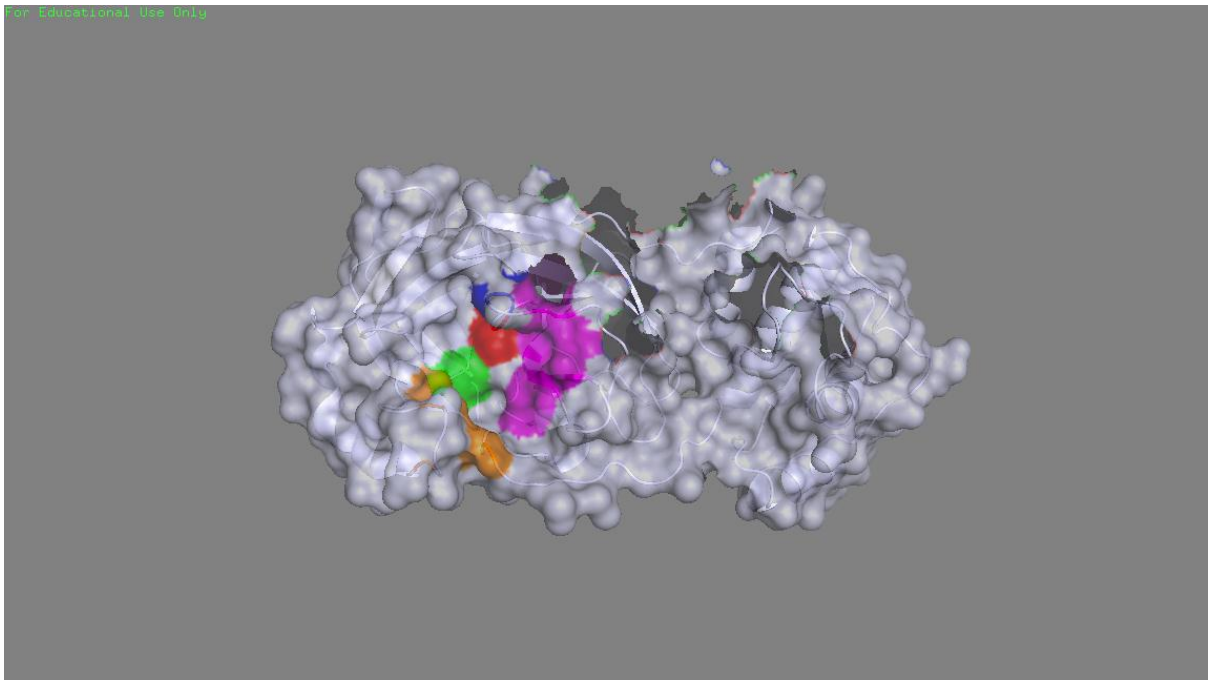


Figure 5.3B- Substrate binding sites on the surface of TGEV main protease:

The surface representation of TGEV main protease is shown in the figure with 40% surface transparency. In the figure, the S1 specificity pocket is shown in magenta color; S2 specificity pocket is shown in orange color and the oxyanion hole is shown in blue color. The active amino acids Cys144 is shown in red; His41 is shown in green (PDB ID-2AMP).



TABLES

Table 1.1- Classification of proteases by IUBMB

Exopeptidases-cleave near the N-or C-termini of peptides and proteins

IUBMB code	**Protease name**
EC 3.4.11	Aminopeptidases
EC 3.4.13	Dipeptides
EC 3.4.14	Dipeptidyl-peptidases and tripeptidyl peptidases
EC 3.4.15	Peptidyl-dipeptidases
EC 3.4.16	Serine-type carboxypeptidases
EC 3.4.17	metallo carboxypeptidases
EC 3.4.18	Cysteine-type carboxypeptidases

Endopeptidases-these peptidases cleave internal peptide bonds in proteins and peptides.

IUBMB code	**Protease name**
EC 3.4.21	Serine endopeptidases
EC 3.4.22	Cysteine endopeptidases
EC 3.4.23	Aspartic endopeptidases
EC 3.4.24	Metallo endopeptidases

Table 1.2- Classification of proteases by MEROPS peptidase database:

Aspartic acid proteases:

FAMILY	SUBFAMILY	EXAMPLE
A1	A1A	Pepsin A(*Homo sapiens*).
	A1B	Nepenthesin (*Nepenthes gracilis*).
A2	A2A	HIV-I retropepsin (human immunodeficiency virus I).
	A2B	Ty3 transposon peptidase (Saccharomyces cerevisiae).
A3	A3A	Cauliflower mosaic virus-type peptidase (cauliflower mosaic virus).

Cysteine proteases:

FAMILY	SUBFAMILY	EXAMPLE
C1	C1A	Papain (*Carica papaya*)
	C1B	Bleomycin hydrolase (*Saccharomyces cerevisiae*)
C2	C2A	Calpain-2 (*Homo sapiens*)
C3	C3A	Poliovirus-type picornain 3C (human

	poliovirus 1)
C3B	Enterovirus picornain 2A (human poliovirus 1)

CHAPTER 6-REFERENCES

1. **Abbenante, G., and D. P. Fairlie**. 2005. Protease inhibitors in the clinic. Med Chem **1**:71–104.
2. **Anand, K., G. J. Palm, J. R. Mesters, S. G. Siddell, J. Ziebuhr, and R. Hilgenfeld**. 2002. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. EMBO J. **21**:3213–3224.
3. **Anand, K., J. Ziebuhr, P. Wadhwani, J. R. Mesters, and R. Hilgenfeld**. 2003. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. Science **300**:1763–1767.
4. **Argos, P.** 1987. A sensitive procedure to compare amino acid sequences. J. Mol. Biol. **193**:385–396.
5. **Bae, I., D. J. Jackwood, D. A. Benfield, L. J. Saif, R. D. Wesley, and H. Hill**. 1991. Differentiation of transmissible gastroenteritis virus from porcine respiratory coronavirus and other antigenically related coronaviruses by using cDNA probes specific for the 5' region of the S glycoprotein gene. J. Clin. Microbiol. **29**:215–218.
6. **Barrett, A. J., and J. K. McDonald**. 1986. Nomenclature: protease, proteinase and peptidase. Biochem. J. **237**:935.
7. **Barrett, A. ., N. D. Rawlings, and J. Woessner**. 2004. The Handbook of Proteolytic Enzymes., 2nd ed. Elsevier Press.
8. **Beaudette,, F. ., and C. . Hudson**. 1937. Cultivation of the virus of infectious bronchitis. J. Am. Vet. Med. Assoc. **90**:51–60.
9. **Bernard, S., and H. Laude**. 1995. Site-specific alteration of transmissible gastroenteritis virus spike protein results in markedly reduced pathogenicity. J. Gen. Virol. **76 (Pt 9)**:2235–2241.
10. **Bertolotti-Ciarlet, A., S. E. Crawford, A. M. Hutson, and M. K. Estes**. 2003. The 3' end of Norwalk virus mRNA contains determinants that regulate the expression and stability of the viral capsid protein VP1: a novel function for the VP2 protein. J. Virol. **77**:11603–11615.
11. **Blakeney, S. J., A. Cahill, and P. A. Reilly**. 2003. Processing of Norwalk virus nonstructural proteins by a 3C-like cysteine proteinase. Virology **308**:216–224.
12. **Blanchard, J. E., N. H. Elowe, C. Huitema, P. D. Fortin, J. D. Cechetto, L. D. Eltis, and E. D. Brown**. 2004. High-throughput screening identifies inhibitors of the SARS coronavirus main proteinase. Chem. Biol. **11**:1445–1453.
13. **Boyer, P. D.** 1971. The enzymes, 3rd ed. Academic Press, New York.
14. **Brandenburg, B., L. Y. Lee, M. Lakadamyali, M. J. Rust, X. Zhuang, and J. M. Hogle**. 2007. Imaging poliovirus entry in live cells. PLoS Biol. **5**:e183.
15. **Brown, D. M., C. T. Cornell, G. P. Tran, J. H. C. Nguyen, and B. L. Semler**. 2005. An authentic 3' noncoding region is necessary for efficient poliovirus replication. J. Virol. **79**:11962–11973.
16. **Brown, D. M., S. E. Kauder, C. T. Cornell, G. M. Jang, V. R. Racaniello, and B. L. Semler**. 2004. Cell-dependent role for the poliovirus 3' noncoding region in positive-strand RNA synthesis. J. Virol. **78**:1344–1351.
17. **BROWN, G. C.** 1952. The influence of chemicals on the propagation of poliomyelitis virus in tissue culture. J. Immunol. **69**:441–450.
18. **Bugg, T. D. H.** 2004. Introduction to enzyme and coenzyme chemistry, 2nd ed. Blackwell publishing.
19. **Chen, S., L. Chen, H. Luo, T. Sun, J. Chen, F. Ye, J. Cai, J. Shen, X. Shen, and H. Jiang**. 2005. Enzymatic activity characterization of SARS coronavirus 3C-like

protease by fluorescence resonance energy transfer technique. Acta Pharmacol. Sin. **26**:99–106.
20. **Clark, M. E., T. Hämmerle, E. Wimmer, and A. Dasgupta**. 1991. Poliovirus proteinase 3C converts an active form of transcription factor IIIC to an inactive form: a mechanism for inhibition of host cell polymerase III transcription by poliovirus. EMBO J. **10**:2941–2947.
21. **Cliver, D. O.** 1997. Virus transmission via food. World Health Stat Q **50**:90–101.
22. **Collett, M. S., J. Neyts, and J. F. Modlin**. 2008. A case for developing antiviral drugs against polio. Antiviral Research **79**:179–187.
23. **De Jesus, N. H.** 2007. Epidemics to eradication: the modern history of poliomyelitis. Virol. J. **4**:70.
24. **Delmas, B., J. Gelfi, R. L'Haridon, L. K. Vogel, H. Sjöström, O. Norén, and H. Laude**. 1992. Aminopeptidase N is a major receptor for the entero-pathogenic coronavirus TGEV. Nature **357**:417–420.
25. **Dorsch-Häsler, K., Y. Yogo, and E. Wimmer**. 1975. Replication of picornaviruses. I. Evidence from in vitro RNA synthesis that poly(A) of the poliovirus genome is genetically coded. J. Virol. **16**:1512–1517.
26. **Eleouet, J. F., D. Rasschaert, P. Lambert, L. Levy, P. Vende, and H. Laude**. 1995. Complete sequence (20 kilobases) of the polyprotein-encoding gene 1 of transmissible gastroenteritis virus. Virology **206**:817–822.
27. **Elston, D. M.** 2009. Update on cutaneous manifestations of infectious diseases. Med. Clin. North Am. **93**:1283–1290.
28. **Estes, M. K., B. V. Prasad, and R. L. Atmar**. 2006. Noroviruses everywhere: has something changed? Curr. Opin. Infect. Dis. **19**:467–474.
29. **Fitzgerald, G. R.** Improving the efficacy of oral TGE vaccination.
30. **Fitzgerald, G. R., and C. . Welter**. 1990. The effect of an oral TGE vaccine on eliminating enzootic TGE virus from a herd of swine. Agri. Pract. Disease Control **11**:25–29.
31. **Fitzgerald, P. M., B. M. McKeever, J. F. VanMiddlesworth, J. P. Springer, J. C. Heimbach, C. T. Leu, W. K. Herber, R. A. Dixon, and P. L. Darke**. 1990. Crystallographic analysis of a complex between human immunodeficiency virus type 1 protease and acetyl-pepstatin at 2.0-A resolution. J. Biol. Chem. **265**:14209–14219.
32. **G.K., S.** 2006. Nitrogen starvation promotes biodegradation of N-heterocyclic compounds in soil. Soil Biology and Biochemistry **38**:2478–2480.
33. **Garwes, D. J., and D. H. Pocock**. 1975. The polypeptide structure of transmissible gastroenteritis virus. J. Gen. Virol. **29**:25–34.
34. **Ghosh, A. K., K. Xi, V. Grum-Tokars, X. Xu, K. Ratia, W. Fu, K. V. Houser, S. C. Baker, M. E. Johnson, and A. D. Mesecar**. 2007. Structure-based design, synthesis, and biological evaluation of peptidomimetic SARS-CoV 3CLpro inhibitors. Bioorg. Med. Chem. Lett. **17**:5876–5880.
35. **Glass, R. I., U. D. Parashar, and M. K. Estes**. 2009. Norovirus gastroenteritis. N. Engl. J. Med. **361**:1776–1785.
36. **Godfrey, T., and S. West**. 1996. Industrial enzymology, 2nd ed. Macmillian publishers Inc, New York.
37. **González, J. M., P. Gomez-Puertas, D. Cavanagh, A. E. Gorbalenya, and L. Enjuanes**. 2003. A comparative sequence analysis to revise the current taxonomy of the family Coronaviridae. Arch. Virol. **148**:2207–2235.
38. **Guan, Y., B. J. Zheng, Y. Q. He, X. L. Liu, Z. X. Zhuang, C. L. Cheung, S. W. Luo, P. H. Li, L. J. Zhang, Y. J. Guan, K. M. Butt, K. L. Wong, K. W. Chan, W. Lim, K. F. Shortridge, K. Y. Yuen, J. S. M. Peiris, and L. L. M. Poon**. 2003.

Isolation and Characterization of Viruses Related to the SARS Coronavirus from Animals in Southern China. Science **302**:276 –278.
39. **Hardy, M. E., T. J. Crone, J. E. Brower, and K. Ettayebi**. 2002. Substrate specificity of the Norwalk virus 3C-like proteinase. Virus Res. **89**:29–39.
40. **Harris, K. S., S. R. Reddigari, M. J. Nicklin, T. Hämmerle, and E. Wimmer**. 1992. Purification and characterization of poliovirus polypeptide 3CD, a proteinase and a precursor for RNA polymerase. J. Virol. **66**:7481–7489.
41. **Hartley, B. S.** 1960. Proteolytic Enzymes. Annual Review of Biochemistry **29**:45–72.
42. **Hata, S., T. Sato, H. Sorimachi, S. Ishiura, and K. Suzuki**. 2000. A simple purification and fluorescent assay method of the poliovirus 3C protease searching for specific inhibitors. J. Virol. Methods **84**:117–126.
43. **He, Y., S. Mueller, P. R. Chipman, C. M. Bator, X. Peng, V. D. Bowman, S. Mukhopadhyay, E. Wimmer, R. J. Kuhn, and M. G. Rossmann**. 2003. Complexes of poliovirus serotypes with their common cellular receptor, CD155. J. Virol. **77**:4827–4835.
44. **Hedstrom, L.** Introduction: Proteases. Chemical Reviews **102**:4429–4430.
45. **Hedstrom, L.** 2002. Serine Protease Mechanism and Specificity. Chem. Rev. **102**:4501–4524.
46. **Hegyi, A., and J. Ziebuhr**. 2002. Conservation of substrate specificities among coronavirus main proteases. J. Gen. Virol. **83**:595–599.
47. **Hibbs, M. S., K. A. Hasty, J. M. Seyer, A. H. Kang, and C. L. Mainardi**. 1985. Biochemical and immunological characterization of the secreted forms of human neutrophil gelatinase. J. Biol. Chem. **260**:2493–2500.
48. **Hoffman, T.** 1974. Food related enzymes. Adv. Chem. Ser **136**:146–185.
49. **Hussey, R. J., L. Coates, R. S. Gill, P. T. Erskine, S.-F. Coker, E. Mitchell, J. B. Cooper, S. Wood, R. Broadbridge, I. N. Clarke, P. R. Lambden, and P. M. Shoolingin-Jordon**. A Structural Study of Norovirus 3C Protease Specificity: Binding of a Designed Active Site-Directed Peptide Inhibitor. Biochemistry **50**:240–249.
50. **Jaulent, A. M., A. S. Fahy, S. R. Knox, J. R. Birtley, N. Roqué-Rosell, S. Curry, and R. J. Leatherbarrow**. 2007. A continuous assay for foot-and-mouth disease virus 3C protease activity. Anal. Biochem. **368**:130–137.
51. **Jewell, D. A., W. Swietnicki, B. M. Dunn, and B. A. Malcolm**. 1992. Hepatitis A virus 3C proteinase substrate specificity. Biochemistry **31**:7862–7869.
52. **Jiang, P., J. A. J. Faase, H. Toyoda, A. Paul, E. Wimmer, and A. E. Gorbalenya**. 2007. Evidence for emergence of diverse polioviruses from C-cluster coxsackie A viruses and implications for global poliovirus eradication. Proc. Natl. Acad. Sci. U.S.A. **104**:9457–9462.
53. **Kadaveru, K., J. Vyas, and M. R. Schiller**. 2008. Viral infection and human disease--insights from minimotifs. Front. Biosci. **13**:6455–6471.
54. **Kao, R. Y., A. P. C. To, L. W. Y. Ng, W. H. W. Tsui, T. S. W. Lee, H.-W. Tsoi, and K.-Y. Yuen**. 2004. Characterization of SARS-CoV main protease and identification of biologically active small molecule inhibitors using a continuous fluorescence-based assay. FEBS Lett. **576**:325–330.
55. **Kapikian, A. Z., R. G. Wyatt, R. Dolin, T. S. Thornhill, A. R. Kalica, and R. M. Chanock**. 1972. Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis. J. Virol. **10**:1075–1081.
56. **Kaplan, G., M. S. Freistadt, and V. R. Racaniello**. 1990. Neutralization of poliovirus by cell receptors expressed in insect cells. J. Virol. **64**:4697–4702.
57. **Kazmierski, W. M., T. P. Kenakin, and K. S. Gudmundsson**. 2006. Peptide, peptidomimetic and small-molecule drug discovery targeting HIV-1 host-cell

attachment and entry through gp120, gp41, CCR5 and CXCR4. Chem Biol Drug Des **67**:13–26.
58. **Kenny, A. J., and A. G. Booth**. 1978. Microvilli: their ultrastructure, enzymology and molecular organization. Essays Biochem. **14**:1–44.
59. **Kew, O. M., R. W. Sutter, E. M. de Gourville, W. R. Dowdle, and M. A. Pallansch**. 2005. Vaccine-derived polioviruses and the endgame strategy for global polio eradication. Annu. Rev. Microbiol. **59**:587–635.
60. **Kim, L., J. Hayes, P. Lewis, A. V. Parwani, K. O. Chang, and L. J. Saif**. 2000. Molecular characterization and pathogenesis of transmissible gastroenteritis coronavirus (TGEV) and porcine respiratory coronavirus (PRCV) field isolates co-circulating in a swine herd. Arch. Virol. **145**:1133–1147.
61. **Kim Y. Green**. 2007. Caliciviridae: The Noroviruses., p. 949–979. *In* Fields Virology, 5th. ed. Lippincott Williams & Wilkins., Philadelphia.
62. **Kishimoto, T., H. Kikutani, Y. Nishizawa, N. Sakaguchi, and Y. Yamamura**. 1979. Involvement of anti-Ig-activated serine protease in the generation of cytoplasmic factor(s) that are responsible for the transmission of Ig-receptor-mediated signals. J. Immunol. **123**:1504–1510.
63. **Kitamura, N., B. L. Semler, P. G. Rothberg, G. R. Larsen, C. J. Adler, A. J. Dorner, E. A. Emini, R. Hanecak, J. J. Lee, S. van der Werf, C. W. Anderson, and E. Wimmer**. 1981. Primary structure, gene organization and polypeptide expression of poliovirus RNA. Nature **291**:547–553.
64. **Koike, S., I. Ise, and A. Nomoto**. 1991. Functional domains of the poliovirus receptor. Proc. Natl. Acad. Sci. U.S.A. **88**:4104–4108.
65. **Kräusslich, H. G., M. J. Nicklin, H. Toyoda, D. Etchison, and E. Wimmer**. 1987. Poliovirus proteinase 2A induces cleavage of eucaryotic initiation factor 4F polypeptide p220. Journal of Virology **61**:2711 –2718.
66. **Kräusslich, H. G., and E. Wimmer**. 1988. Viral proteinases. Annu. Rev. Biochem. **57**:701–754.
67. **Kuyumcu-Martinez, M., G. Belliot, S. V. Sosnovtsev, K.-O. Chang, K. Y. Green, and R. E. Lloyd**. 2004. Calicivirus 3C-like proteinase inhibits cellular translation by cleavage of poly(A)-binding protein. J. Virol. **78**:8172–8182.
68. **Labbé, J. P., P. Rebeyrotte, and M. Turpin**. 1974. [Demonstration of extracellular leucine aminopeptidases (EC 3.4.1) of Aspergillus oryzae (IP 410). Study of the leucine aminopeptidase 2 fraction]. C.R. Hebd. Seances Acad. Sci., Ser. D, Sci. Nat. **278**:2699–2702.
69. **Lambden, P. R., E. O. Caul, C. R. Ashley, and I. N. Clarke**. 1993. Sequence and genome organization of a human small round-structured (Norwalk-like) virus. Science **259**:516–519.
70. **Lambden, P. R., B. Liu, and I. N. Clarke**. 1995. A conserved sequence motif at the 5' terminus of the Southampton virus genome is characteristic of the Caliciviridae. Virus Genes **10**:149–152.
71. **Lawson, M. A., and B. L. Semler**. 1992. Alternate poliovirus nonstructural protein processing cascades generated by primary sites of 3C proteinase cleavage. Virology **191**:309–320.
72. **Leblanc, P., E. Pattou, A. L'heritier, and C. Kordon**. 1980. Some properties of peptidasic activity bound to the anterior pituitary membranes. Biochemical and Biophysical Research Communications **96**:1457–1465.
73. **Lindesmith, L., C. Moe, S. Marionneau, N. Ruvoen, X. Jiang, L. Lindblad, P. Stewart, J. LePendu, and R. Baric**. 2003. Human susceptibility and resistance to Norwalk virus infection. Nat. Med. **9**:548–553.

74. **M. C. Lai, M., S. Perlman, and L. J. Anderson**. 2007. Coronaviridae, p. 1305–1335. *In* Fields Virology, 5th ed. Lippincott Williams & Wilkins., Philadelphia.
75. **Madala, P. K., J. D. A. Tyndall, T. Nall, and D. P. Fairlie**. 2010. Update 1 of: Proteases universally recognize beta strands in their active sites. Chem. Rev. **110**:PR1–31.
76. **Marra, M. A., S. J. M. Jones, C. R. Astell, R. A. Holt, A. Brooks-Wilson, Y. S. N. Butterfield, J. Khattra, J. K. Asano, S. A. Barber, S. Y. Chan, A. Cloutier, S. M. Coughlin, D. Freeman, N. Girn, O. L. Griffith, S. R. Leach, M. Mayo, H. McDonald, S. B. Montgomery, P. K. Pandoh, A. S. Petrescu, A. G. Robertson, J. E. Schein, A. Siddiqui, D. E. Smailus, J. M. Stott, G. S. Yang, F. Plummer, A. Andonov, H. Artsob, N. Bastien, K. Bernard, T. F. Booth, D. Bowness, M. Czub, M. Drebot, L. Fernando, R. Flick, M. Garbutt, M. Gray, A. Grolla, S. Jones, H. Feldmann, A. Meyers, A. Kabani, Y. Li, S. Normand, U. Stroher, G. A. Tipples, S. Tyler, R. Vogrig, D. Ward, B. Watson, R. C. Brunham, M. Krajden, M. Petric, D. M. Skowronski, C. Upton, and R. L. Roper**. 2003. The Genome sequence of the SARS-associated coronavirus. Science **300**:1399–1404.
77. **Martín, J.** 2006. Vaccine-derived poliovirus from long term excretors and the end game of polio eradication. Biologicals **34**:117–122.
78. **Matsui, S. M., and H. B. Greenberg**. 2000. Immunity to calicivirus infection. J. Infect. Dis. **181 Suppl 2**:S331–335.
79. **Mendelsohn, C. L., E. Wimmer, and V. R. Racaniello**. 1989. Cellular receptor for poliovirus: molecular cloning, nucleotide sequence, and expression of a new member of the immunoglobulin superfamily. Cell **56**:855–865.
80. **Menon, A. S., and A. L. Goldberg**. 1987. Protein substrates activate the ATP-dependent protease La by promoting nucleotide binding and release of bound ADP. J. Biol. Chem. **262**:14929–14934.
81. **Mitchell, R. S., V. Kumar, A. K. Abbas, and N. Fausto**. 2007. Robins Basic Pathology, 8th ed. Saunders, Philadelphia.
82. **Mosimann, S. C., M. M. Cherney, S. Sia, S. Plotch, and M. N. . James**. 1997. Refined X-ray crystallographic structure of the poliovirus 3C gene product. Journal of Molecular Biology **273**:1032–1047.
83. **Mueller, S., and E. Wimmer**. 2003. Recruitment of nectin-3 to cell-cell junctions through trans-heterophilic interaction with CD155, a vitronectin and poliovirus receptor that localizes to alpha(v)beta3 integrin-containing membrane microdomains. J. Biol. Chem. **278**:31251–31260.
84. **Mueller, S., E. Wimmer, and J. Cello**. 2005. Poliovirus and poliomyelitis: a tale of guts, brains, and an accidental event. Virus Res. **111**:175–193.
85. **Nakamura, K., Y. Someya, T. Kumasaka, G. Ueno, M. Yamamoto, T. Sato, N. Takeda, T. Miyamura, and N. Tanaka**. 2005. A norovirus protease structure provides insights into active and substrate binding site integrity. J. Virol. **79**:13685–13693.
86. **Nathanson, N., and O. M. Kew**. 2010. From emergence to eradication: the epidemiology of poliomyelitis deconstructed. Am. J. Epidemiol. **172**:1213–1229.
87. **Nduwimana, J., L. Guenet, I. Dorval, M. Blayau, J. Y. Le Gall, and A. Le Treut**. 1995. Proteases. Ann. Biol. Clin. (Paris) **53**:251–264.
88. **Oberste, M. S., K. Maher, M. R. Flemister, G. Marchetti, D. R. Kilpatrick, and M. A. Pallansch**. 2000. Comparison of classic and molecular approaches for the identification of untypeable enteroviruses. J. Clin. Microbiol. **38**:1170–1174.

89. **Ohka, S., W. X. Yang, E. Terada, K. Iwasaki, and A. Nomoto**. 1998. Retrograde transport of intact poliovirus through the axon via the fast transport system. Virology **250**:67–75.
90. **Okada, Y., H. Nagase, and E. D. Harris Jr**. 1986. A metalloproteinase from human rheumatoid synovial fibroblasts that digests connective tissue matrix components. Purification and characterization. J. Biol. Chem. **261**:14245–14255.
91. **Oliver, S. L., E. Asobayire, A. M. Dastjerdi, and J. C. Bridger**. 2006. Genomic characterization of the unclassified bovine enteric virus Newbury agent-1 (Newbury1) endorses a new genus in the family Caliciviridae. Virology **350**:240–250.
92. **Pallansch, M., and R. Ross**. 2007. Enteroviruses: Polioviruses, Coxsackieviruses, Echoviruses, and Newer Enteroviruses., p. 839–893. *In* Fields Virology, 5th ed. Lippincott Williams & Wilkins., Philadelphia.
93. **Patel, M. M., A. J. Hall, J. Vinjé, and U. D. Parashar**. 2009. Noroviruses: A comprehensive review. Journal of Clinical Virology **44**:1–8.
94. **Paul JR.** 1971. A History of Poliomyelitis. Yale studies in the history of science and medicine., 1st ed. Yale University Press, New Haven & London.
95. **Pensaert, M., P. Callebaut, and J. Vergote**. 1986. Isolation of a porcine respiratory, non-enteric coronavirus related to transmissible gastroenteritis. Vet Q **8**:257–261.
96. **Pontremoli, S., F. Salamino, B. Sparatore, E. Melloni, A. Morelli, U. Benatti, and A. De Flora**. 1979. Isolation and partial characterization of three acidic proteinases in erythrocyte membranes. Biochem. J. **181**:559–568.
97. **Powers, J. C., J. L. Asgian, O. D. Ekici, and K. E. James**. 2002. Irreversible inhibitors of serine, cysteine, and threonine proteases. Chem. Rev. **102**:4639–4750.
98. **Pringle, C. R.** 1999. Virus Taxonomy - 1999. Archives of Virology **144**:421–429.
99. **Racaniello, V. R., and D. Baltimore**. 1981. Cloned poliovirus complementary DNA is infectious in mammalian cells. Science **214**:916–919.
100. **Racaniello, V. R.** 2006. One hundred years of poliovirus pathogenesis. Virology **344**:9–16.
101. **Racaniello, V. R.** 2007. Picornaviridae: The Viruses and Their Replication., p. 795–838. *In* Fields Virology, 5th ed. Lippincott Williams & Wilkins., Philadelphia.
102. **Ramirez, S., G. M. Giammanco, S. De Grazia, C. Colomba, V. Martella, and S. Arista**. 2008. Genotyping of GII.4 and GIIb norovirus RT-PCR amplicons by RFLP analysis. J. Virol. Methods **147**:250–256.
103. **Rankin, B. B., T. M. McIntyre, and N. P. Curthoys**. 1980. Brush border membrane hydrolysis of S-benzyl-cysteine-p-nitroanilide, and activity of aminopeptidase M. Biochem. Biophys. Res. Commun. **96**:991–996.
104. **Rao, M. B., A. M. Tanksale, M. S. Ghatge, and V. V. Deshpande**. 1998. Molecular and Biotechnological Aspects of Microbial Proteases. Microbiology and Molecular Biology Reviews **62**:597 –635.
105. **Rasschaert, D., M. Duarte, and H. Laude**. 1990. Porcine respiratory coronavirus differs from transmissible gastroenteritis virus by a few genomic deletions. Journal of General Virology **71**:2599 –2607.
106. **Rawlings, N. D., and A. J. Barrett**. 1995. Evolutionary families of metallopeptidases. Meth. Enzymol. **248**:183–228.
107. **Rawlings, N. D., and A. J. Barrett**. 1993. Evolutionary families of peptidases. Biochem J **290**:205–218.
108. **Rawlings, N. D., and A. J. Barrett**. 1995. Families of aspartic peptidases, and those of unknown catalytic mechanism. Meth. Enzymol. **248**:105–120.
109. **Rawlings, N. D., and A. J. Barrett**. 1994. Families of cysteine peptidases. Meth. Enzymol. **244**:461–486.

110. **Rawlings, N. D., and A. J. Barrett**. 1994. Families of serine peptidases. Meth. Enzymol. **244**:19–61.
111. **Rawlings, N. D., A. J. Barrett, and A. Bateman**. 2010. MEROPS: the peptidase database. Nucleic Acids Res. **38**:D227–233.
112. **Rawlings, N. D., F. R. Morton, C. Y. Kok, J. Kong, and A. J. Barrett**. 2008. MEROPS: the peptidase database. Nucleic Acids Res. **36**:D320–325.
113. **Reeves, J. D., and A. J. Piefer**. 2005. Emerging drug targets for antiretroviral therapy. Drugs **65**:1747–1766.
114. **Ren, R., and V. R. Racaniello**. 1992. Poliovirus spreads from muscle to the central nervous system by neural pathways. J. Infect. Dis. **166**:747–752.
115. **Roberts, A., L. Vogel, J. Guarner, N. Hayes, B. Murphy, S. Zaki, and K. Subbarao**. 2005. Severe acute respiratory syndrome coronavirus infection of golden Syrian hamsters. J. Virol. **79**:503–511.
116. **Ronald Ross, W.** 1976. Chapter I Substrate Specificities of Aminopeptidases: A Specific Method for Microbial Differentiation, p. 1–14. *In* Methods in Microbiology. Academic Press.
117. **Rota, P. A., M. S. Oberste, S. S. Monroe, W. A. Nix, R. Campagnoli, J. P. Icenogle, S. Peñaranda, B. Bankamp, K. Maher, M.-H. Chen, S. Tong, A. Tamin, L. Lowe, M. Frace, J. L. DeRisi, Q. Chen, D. Wang, D. D. Erdman, T. C. T. Peret, C. Burns, T. G. Ksiazek, P. E. Rollin, A. Sanchez, S. Liffick, B. Holloway, J. Limor, K. McCaustland, M. Olsen-Rasmussen, R. Fouchier, S. Günther, A. D. M. E. Osterhaus, C. Drosten, M. A. Pallansch, L. J. Anderson, and W. J. Bellini**. 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science **300**:1394–1399.
118. **SABIN, A. B.** 1956. Pathogenesis of poliomyelitis; reappraisal in the light of new data. Science **123**:1151–1157.
119. **Sair, A. I., D. H. D'Souza, and L. A. Jaykus**. 2002. Human Enteric Viruses as Causes of Foodborne Disease. Comprehensive Reviews in Food Science and Food Safety **1**:73–89.
120. **Sárkány, Z., and L. Polgár**. 2003. The unusual catalytic triad of poliovirus protease 3C. Biochemistry **42**:516–522.
121. **Schechter, I., and A. Berger**. 1967. On the size of the active site in proteases. I. Papain. Biochemical and Biophysical Research Communications **27**:157–162.
122. **Scheffler, U., W. Rudolph, J. Gebhardt, and J. Rohayem**. 2007. Differential cleavage of the norovirus polyprotein precursor by two active forms of the viral protease. J. Gen. Virol. **88**:2013–2018.
123. **Schultze, B., C. Krempl, M. L. Ballesteros, L. Shaw, R. Schauer, L. Enjuanes, and G. Herrler**. 1996. Transmissible gastroenteritis coronavirus, but not the related porcine respiratory coronavirus, has a sialic acid (N-glycolylneuraminic acid) binding activity. J. Virol. **70**:5634–5637.
124. **Schwegmann-Wessels, C., and G. Herrler**. 2006. Sialic acids as receptor determinants for coronaviruses. Glycoconj. J. **23**:51–58.
125. **Shannon, J. D., E. N. Baramova, J. B. Bjarnason, and J. W. Fox**. 1989. Amino acid sequence of a Crotalus atrox venom metalloproteinase which cleaves type IV collagen and gelatin. J. Biol. Chem. **264**:11575–11583.
126. **Siebenga, J. J., H. Vennema, D.-P. Zheng, J. Vinjé, B. E. Lee, X.-L. Pang, E. C. M. Ho, W. Lim, A. Choudekar, S. Broor, T. Halperin, N. B. G. Rasool, J. Hewitt, G. E. Greening, M. Jin, Z.-J. Duan, Y. Lucero, M. O'Ryan, M. Hoehne, E. Schreier, R. M. Ratcliff, P. A. White, N. Iritani, G. Reuter, and M. Koopmans**. 2009.

Norovirus illness is a global problem: emergence and spread of norovirus GII.4 variants, 2001-2007. J. Infect. Dis. **200**:802–812.
127. **Someya, Y., and N. Takeda**. 2009. Insights into the enzyme-substrate interaction in the norovirus 3C-like protease. J. Biochem. **146**:509–521.
128. **Someya, Y., N. Takeda, and T. Wakita**. 2008. Saturation mutagenesis reveals that GLU54 of norovirus 3C-like protease is not essential for the proteolytic activity. J. Biochem. **144**:771–780.
129. **Spector, D. H., and D. Baltimore**. 1974. Requirement of 3'-terminal poly(adenylic acid) for the infectivity of poliovirus RNA. Proc. Natl. Acad. Sci. U.S.A. **71**:2983–2987.
130. **Straus, J. W., R. F. Parrish, and K. L. Polakoski**. 1981. Boar acrosin. Association of an endogenous membrane proteinase with phospholipid membranes. J. Biol. Chem. **256**:5662–5668.
131. **Strebel, P. M., R. W. Sutter, S. L. Cochi, R. J. Biellik, E. W. Brink, O. M. Kew, M. A. Pallansch, W. A. Orenstein, and A. R. Hinman**. 1992. Epidemiology of poliomyelitis in the United States one decade after the last reported case of indigenous wild virus-associated disease. Clin. Infect. Dis. **14**:568–579.
132. **Tan, M., M. Jin, H. Xie, Z. Duan, X. Jiang, and Z. Fang**. 2008. Outbreak studies of a GII-3 and a GII-4 norovirus revealed an association between HBGA phenotypes and viral infection. J. Med. Virol. **80**:1296–1301.
133. **Tong, L.** 2002. Viral proteases. Chem. Rev. **102**:4609–4626.
134. **Tsantrizos, Y. S.** 2008. Peptidomimetic therapeutic agents targeting the protease enzyme of the human immunodeficiency virus and hepatitis C virus. Acc. Chem. Res. **41**:1252–1263.
135. **van der Hoek, L., K. Pyrc, M. F. Jebbink, W. Vermeulen-Oost, R. J. M. Berkhout, K. C. Wolthers, P. M. E. Wertheim-van Dillen, J. Kaandorp, J. Spaargaren, and B. Berkhout**. 2004. Identification of a new human coronavirus. Nat. Med. **10**:368–373.
136. **van der Hoorn, R. A. L.** 2008. Plant proteases: from phenotypes to molecular mechanisms. Annu Rev Plant Biol **59**:191–223.
137. **Weaver, L. H., W. R. Kester, and B. W. Matthews**. 1977. A crystallographic study of the complex of phosphoramidon with thermolysin. A model for the presumed catalytic transition state and for the binding of extended substances. J. Mol. Biol. **114**:119–132.
138. **Weidner, J. R., and B. M. Dunn**. 1991. Development of synthetic peptide substrates for the poliovirus 3C proteinase. Arch. Biochem. Biophys. **286**:402–408.
139. **Widdowson, M.-A., A. Sulka, S. N. Bulens, R. S. Beard, S. S. Chaves, R. Hammond, E. D. P. Salehi, E. Swanson, J. Totaro, R. Woron, P. S. Mead, J. S. Bresee, S. S. Monroe, and R. I. Glass**. 2005. Norovirus and foodborne disease, United States, 1991-2000. Emerging Infect. Dis. **11**:95–102.
140. **Wimmer, E., C. U. T. Hellen, and X. Cao**. 1993. Genetics of Poliovirus. Annual Review of Genetics **27**:353–436.
141. **Wolfe, M. S.** 2009. Intramembrane proteolysis. Chem. Rev. **109**:1599–1612.
142. **Woo, P. C. Y., S. K. P. Lau, C. Chu, K. Chan, H. Tsoi, Y. Huang, B. H. L. Wong, R. W. S. Poon, J. J. Cai, W. Luk, L. L. M. Poon, S. S. Y. Wong, Y. Guan, J. S. M. Peiris, and K. Yuen**. 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. J. Virol. **79**:884–895.
143. **Xi, J. N., D. Y. Graham, K. N. Wang, and M. K. Estes**. 1990. Norwalk virus genome cloning and characterization. Science **250**:1580–1583.

144. **Xue, X., H. Yu, H. Yang, F. Xue, Z. Wu, W. Shen, J. Li, Z. Zhou, Y. Ding, Q. Zhao, X. C. Zhang, M. Liao, M. Bartlam, and Z. Rao**. 2008. Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. J. Virol. **82**:2515–2527.
145. **Yafal, A. G., G. Kaplan, V. R. Racaniello, and J. M. Hogle**. 1993. Characterization of Poliovirus Conformational Alteration Mediated by Soluble Cell Receptors. Virology **197**:501–505.
146. **Yang, C.-W., Y.-N. Yang, P.-H. Liang, C.-M. Chen, W.-L. Chen, H.-Y. Chang, Y.-S. Chao, and S.-J. Lee**. 2007. Novel small-molecule inhibitors of transmissible gastroenteritis virus. Antimicrob. Agents Chemother. **51**:3924–3931.
147. **Yang, H., W. Xie, X. Xue, K. Yang, J. Ma, W. Liang, Q. Zhao, Z. Zhou, D. Pei, J. Ziebuhr, R. Hilgenfeld, K. Y. Yuen, L. Wong, G. Gao, S. Chen, Z. Chen, D. Ma, M. Bartlam, and Z. Rao**. 2005. Design of wide-spectrum inhibitors targeting coronavirus main proteases. PLoS Biol. **3**:e324.
148. **Yang, H., M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, G. F. Gao, K. Anand, M. Bartlam, R. Hilgenfeld, and Z. Rao**. 2003. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. Proc. Natl. Acad. Sci. U.S.A. **100**:13190–13195.
149. **Yang, W. X., T. Terasaki, K. Shiroki, S. Ohka, J. Aoki, S. Tanabe, T. Nomura, E. Terada, Y. Sugiyama, and A. Nomoto**. 1997. Efficient delivery of circulating poliovirus to the central nervous system independently of poliovirus receptor. Virology **229**:421–428.
150. **Zeitler, C. E., M. K. Estes, and B. V. Venkataram Prasad**. 2006. X-ray crystallographic structure of the Norwalk virus protease at 1.5-A resolution. J. Virol. **80**:5050–5058.
151. **Ziebuhr, J., E. J. Snijder, and A. E. Gorbalenya**. 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. J. Gen. Virol. **81**:853–879.
152. Department of Health and Ageing Norovirus laboratory case definition.
153. 1992. Enzyme Nomenclature. "Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes. Academic Press, Orlando.
154. Morbidity and Mortality Weekly Report (MMWR). Updated Norovirus Outbreak Management and Disease Prevention Guidelines. CDC.
155. Norovirus: Technical Fact Sheet. National center for Infectious Diseases. CDC.
156. The pigsite.com. Porcine respiratory coronavirus infection (PRCoV). Cited on January 2012.