Introduction
INTRODUCTION

1.1 GENERAL

Proteolytic enzymes, also known as proteases/peptidases, are enzymes that catalyze the breakdown of proteins by hydrolysis of peptide bonds. Proteinases are essential for the survival of all kinds of organisms, and are encoded by approximately 2% of all genes [Rawlings et al., 2004b]. Proteases are customarily classified as exopeptidases when they hydrolyze only the N- or C- terminal bonds in proteins and endopeptidases (proteinases) when they hydrolyze internal peptide bonds. However, proteases and proteinases are synonymously used in literature. Based on the catalytic mechanism, there are four types of proteases, serine, cysteine, threonine or aspartic proteases and metallo proteases. There is a recent report on glutamic proteases, being the only subtype not found in mammals so far. The primary role of proteinases was long considered to be protein degradation relevant to food digestion and intracellular protein turnover [Barrett et al., 2004]. However, now it is known that proteinases are involved in control of large number of key physiological processes such as cell-cycle progression, cell proliferation and cell death, DNA replication, tissue remodelling, haemostasis (coagulation), wound healing and immune response [Turk, 2006].

Cysteine proteinases (CPs) are the proteins with molecular mass about 21-30 kDa, showing the highest hydrolytic activity at pH 4-6.5. CPs are present in all living organisms. They are synthesized in a precursor form in order to prevent unwanted proteolysis and later subjected to cotranslational and posttranslational modifications to convert them into catalytically active mature enzymes [Turk et al., 2000]. Till now, 21 families of CPs have been discovered, almost half of them in viruses and rest of them in bacteria, protozoa, fungi, plants and mammals [Barrett et al., 2004; Barrett et al., 2001]. The first clearly recognized and extensively investigated cysteine proteinase is papain isolated from the latex of plant Carica papaya. Mammalian CPs are divided into 4 main groups namely:

1. Lysosomal cathepsins
2. Caspases
3. Calpains
4. Legumains

Cathepsins comprise an important section of the papain family of CPs, sharing similar amino acid sequences and folds. There are eleven human cathepsins known at the sequence level [Turk et al., 2001; Rossi et al., 2004]. Out of which seven viz.
cathepsins B, H, L, C, O, F and X are ubiquitous, such that they have a broad tissue
distribution, but they may be involved in more specialized processes [Turk et al.,
2000; Buhling et al., 2000]. Cathepsins K, V and S are more tissue specific with
cathepsin K expressed in osteoclasts only, cathepsin V in thymus and testis, and
cathepsin S in spleen and lung [Turk et al., 2000; Buhling et al., 2000]. Recently,
cathepsin K was found to be expressed by breast carcinoma cell, mature macrophages,
and multinucleate giant cells adjacent to amyloid deposits in brain [Punturieri et al.,
2000; Zaidi et al., 2001; Rocken et al., 2001]. Cathepsins are all relatively small
monomeric proteins with molecular mass (Mr) in the range of 24-35 kDa, with
exception of cathepsin C, which is an oligomeric enzyme with Mr around 200 kDa
[Turk et al., 2002]. All mature cathepsins are glycosylated at usually one or more
glycosylation sites except cathepsin S. Human cathepsins play very important role in
intracellular protein turnover in lysosomes, processing and activation of other proteins
including proteinases, antigen processing and presentation and in bone remodelling.
However, their specific and individual functions are often associated with their
restricted tissue localization [Brix et al., 2008]. Lysosomal CPs have been found to
be critical for rheumatoid arthritis, osteoarthritis and osteoporosis [Vasiljeva et al.,
2007], neurological disorders [Nakanishi, 2003], pancreatitis [van Acker et al., 2002],
cancer [Kepler, 2006; Goecheva and Joyce, 2007], cardiovascular diseases [Lutgens
et al., 2007], etc. Cathepsins also participate in apoptosis, although there role is still
not clear [Stoka et al., 2005; Turk and Stoka, 2007]. In some pathological conditions
like ischemia, hypervitaminosis and exposure to UV radiations lysosomal enzymes
are released in extracellular space and produce extensive damage to the extracellular
matrix.

**Calpains and Caspases** are cytoplasmic thiol proteinases. Calpains participate in
many intracellular processes like turnover of cytoskeletal proteins, cell differentiation
and regulation of signal peptides. They require Ca$^{2+}$ for activation. They are
ubiquitously distributed and have been implicated in acute neurological disorders,
Alzheimer's disease, muscular dystrophy and gastric cancer [Huang and Wang, 2001].
Caspases are cysteine dependent aspartate specific proteinases. They are involved in
cytokine maturation, apoptosis signalling and mediation [Goyal, 2001].

**Legumains** are cysteine-dependent asparagine endopeptidases. They are involved in
MHC class II-restricted antigen presentation [Manoury et al., 1998] and local
negative regulation of osteoclasts formation and activity [Choi et al., 1994].

**REGULATION OF LYSOSOMAL THIOL PROTEINASE ACTIVITY**

Despite their life-giving functions, the enormous hydrolytic potential of cathepsins can be damaging in living systems and needs to be kept strictly under control. Failures in biological mechanisms controlling proteinase activities result in many diseases such as neurodegeneration, cardiovascular diseases, osteoporosis, arthritis and cancer. Cells have evolved several distinct mechanisms for the regulation of excessive CP activity via proper gene transcription, maintenance of the rate of proteinase synthesis and degradation and most importantly the interaction of CPs with the proteins that inhibit them, viz. cysteine proteinase inhibitors or thiol proteinase inhibitors (CPIs or TPIs) or more commonly cystatins.

The cystatin superfamily comprises a large group of the cystatin domain containing proteins present in wide variety of organisms including humans. Cystatin inhibitory activity is vital for the regulation of normal physiological processes by limiting the potentially inappropriate activity of their target proteinases, cathepsins, mammalian legumain and some calpains [Alvarez-Fernandez et al., 1999; Crawford, 1987].

**1.2 DISCOVERY OF THE CYSTATIN SUPERFAMILY**

Hayashi et al., [1960] for the first time reported the presence of a factor capable of inhibiting the clotting activity of a thiol proteinase in mammalian system. The first isolated and partially characterized protein inhibitor of CPs was from chicken egg white and was shown to inhibit papain, ficin [Fossum and Whitaker, 1968; Sen and Whitaker, 1973] and cathepsins B and C [Keilova and Tomasek, 1975]. Later for the same protein term cystatin was proposed because of its unique property of arresting the activity of CPs [Barrett, 1981]. The first intracellular protein inhibitor of papain, cathepsin B and H was isolated and partially characterized from pig leucocytes and spleen [Kopitar et al., 1978]. The determined amino acid sequences of chicken cystatin [Turk et al., 1983; Schwabe et al., 1984] and human stefin (stefin A) from the cytosol of polymorphonuclear granulocytes [Machleidt et al., 1983] confirmed structural differences between these two homologous proteins. At the same time, inhibitors of CPs were isolated from sera of patients suffering from autoimmune diseases [Turk et al., 1983] and based on its sequence homology with chicken cystatin the name human cystatin was proposed [Brzin et al., 1984], soon renamed to human
cystatin C (HCC) [Barrett et al., 1984]. Sequences of bovine and human kininogens were determined [Nawa et al., 1983; Muller-Esterl et al., 1985a]. And concept of cystatin “superfamily” emerged precipitated by an observation that multiple cystatin-like sequences were present in kininogens and that stefins were related to both cystatins and repeats of kininogens [Ohkubo et al., 1984]. This data and First International Symposium on Cysteine Proteinases and their Inhibitors (Portoroz, Yugoslavia (now Slovenia) September 1985, organized by V. Turk) were crucial for nomenclature and classification of the cystatin superfamily [Barrett et al., 1986].

1.3 CLASSIFICATION OF THE CYSTATIN SUPERFAMILY

The first classification of the cystatin superfamily into three families was based on at least 50% sequence identity, inhibition of their target enzymes and presence or absence of disulphide bonds [Barrett et al., 1986]. Three distinct families of the protein inhibitors comprise: family 1 or the stefin family, family 2 or the cystatin family and family 3 or the kininogen family. The first two families are single domain inhibitors whereas the kininogens are composed of three domains, two being inhibitory [Fig. 1]. A typical cystatin domain was defined to be an approximately 100-amino acids polypeptide that folds into a five stranded β-sheet, which partially wraps around a central α-helix [Fig. 2] [Bode et al., 1988]. Later, the term ‘type’ was introduced and the mammalian cystatins were divided into types 1, 2 and 3 [Rawlings and Barrett, 1990]. However, an increasing number of cystatins from various sources introduced new subdivision of the cystatins into four families [Rawlings and Barrett, 1990], the fourth family consisting of non-inhibitory homologues of two cystatin-like domains, such as human α2 SH-glycoprotein (feutin) and histidine-rich glycoprotein [Brown and Dziegielewska, 1997]. The cystatin superfamily also comprises of phytocystatins. According to recently proposed classification of peptidase inhibitors into families and clans [Rawlings et al., 2004a] cystatins are assigned to family I25 which consists of three subfamilies, I25A (stefins), I25B (cystatins), I25C (are mostly not proteinase inhibitors).
Fig. 1 A diagrammatic representation of the chain structures of proteins in the cystatin superfamily

The stefins are single chain proteins without disulphide linkages. The cystatins are also single chain but possess two disulphide bonds. The structure indicated for the kininogens is that of L- and T-kininogen; H-kininogens have a longer carboxyl terminal extension. There is an additional disulphide link from segment 1 to the kinin segment. The symbol marks potential sites for the attachment of the carbohydrate side chains.
Type I (Stefins)

Type II (Cystatins)

Type III (Kininogens)

Segment 1 | Segment 2 | Segment 3

Kinin
Fig. 2  **Fold of cystatin C**

Cystatin C chain trace is shown in green in orientation which positions the N-terminal "elephant trunk" and the first and the second hairpin loops to the bottom from left to right [Adapted from Bode et al. 1988, EMBO J, 2593-2599].
1.4 GENERAL PROPERTIES OF CYSTATIN SUPERFAMILY

TYPE 1 CYSTATINS (STEFINS)

The protein inhibitors belonging to stefin family are acidic single chain proteins lacking disulphide bonds and carbohydrates, composed of ~100 amino acid residues with Mr of 11 kDa. They are primarily intracellular cytoplasmic proteins of many cell types, although they have been found in extracellular fluids as well [Abrahamson et al., 1986]. In mammals, including human, rat, bovine, mouse and porcine, two members of the stefin family, stefin A (cystatin A or α) and stefin B (cystatin B or β) have been identified [Turk et al., 1997; Barrett et al., 1986]. In addition stefin C was discovered in bovine thymus as the first tryptophan containing stefin with a prolonged N-terminus [Turk et al., 1993] and stefin D in pigs [Lenarcic et al., 1996]. At least three different stefin A variants are encoded within the mouse genome [Tsui et al., 1993]. The type 1 cystatins belong to the subfamily I25A [Rawlings et al., 2004a].

Cystatin A

It is an inhibitor of cathepsin B in human skin discovered by Fraki [1976]. Later on Jarvinen [1978] studied it as ‘acid cysteine proteinase inhibitor’ (ACPI) because of its acidic pI at 4.7-5.0. Brzin et al. [1983] purified an inhibitor from blood leucocytes, and named it as “stefin’. The amino acid sequence was determined by Machleidt et al. [1983]. Green et al. [1984] characterized same type of CPI from human liver and latter renamed it as cystatin A. Cystatin A occurs in multiple isoelectric forms with predominantly acidic pI values in the range 4.5-5.0 [Hopsu-Havu et al., 1983a]. Rinnie et al. [1978] detected cystatin A in extracts of squamous epithelia from oesophagus. It was also found in dendritic reticulum cells of the lymph nodes [Rinnie et al., 1983], seminal plasma [Minakata and Asano, 1985], saliva, bovine skin [Turk et al., 1995] and in a number of epidermoid carcinomas [Rinnie et al., 1984].

Cystatin α is assumed to be a species variant of cystatin A found in rats. This protein was characterized by Jarvinen [1976] as a specific inhibitor of CP from rat skin having Mr of 13 kDa. Cystatin α is generally found on the epidermal layer [Jarvinen et al., 1978] and various other squamous epithelia [Rinnie et al., 1978]. Human stefin A is expressed at high levels in skin and presumably controls cysteine proteinases in the skin. Some cathepsins play a crucial role in the antigen presentation process indicating that the interactions between stefin A and cathepsins contribute to
the species dependent diversity of the endosomal compartments which participate in the immune response [Mihelic et al., 2006].

**Cystatin B**

Cystatin B was detected as an inhibitor of cathepsin B and H in human tissues by Lenney et al. [1979]. It has been purified from human spleen and liver [Jarvinen and Rinnie, 1982; Green et al., 1984]. Cystatin B is relatively basic protein with pI values of 6.25 and 6.35 for its two forms [Green et al., 1984]. It forms dimer [Green et al., 1984] which shows no inhibitory activity. With ubiquitous distribution it appears to be general inhibitor in the cytoplasm. **Cystatin β** a species variant of cystatin B was isolated from rat liver [Finkelstadt, 1957; Lenney, 1979] with pI values ranging from 5.04 to 5.6 [Kominami et al., 1981]. It has even distribution in tissues and is more abundant than cystatin α in all tissues except skin. In this it resembles cystatin B of human variant.

**Stefin C**

Stefin C is unique among the inhibitors from stefin family which was found in multiple forms resulting from the cleavage of Asn 5-Leu 6 bond of the inhibitor. Its MW is calculated to be 11,546 (101 amino acid residues). It was found to be acidic with pI values from 4.5 to 5.6 [Turk et al., 1993].

**TYPE 2 CYSTATINS (CYSTATINS)**

Cystatins typically comprise ~115 amino acids (Mr ~13-15 kDa), are largely acidic (with exception of human cystatin C) contain four conserved cysteine residues known to form two disulphide bonds [Grubb et al., 1984], usually non-glycosylated with exceptions of cystatin E/M [Ni et al., 1997; Sotirpoulou et al., 1997], cystatin F [Ni et al., 1998] and cystatin S [Esnard et al., 1990] which are glycoproteins. They are synthesized with 20-26 residues long signal peptides and are mainly extracellular, secreted proteins, occurring at relatively high concentrations in body fluids [Abrahamson et al., 1986; Kopitar-Jerala, 2006]. Similar to stefins, the cystatins contain the conserved QXVXG region in the central part of the molecule and the P-W pair in the C-terminal part of cystatins [Turk and Bode, 1991]. Chicken cystatin and HCC represent founding members of this family [Turk and Bode, 1991; Abrahamson et al., 1986]. Human type cystatins include cystatin C, D, S, SA and N with about
50% or less sequence identity [Turk and Bode, 1991; Abrahamson et al., 1986; Balbin et al., 1994]. The human type 2 cystatins are grouped in subfamily 125B of the cystatin family [Rawlings et al., 2004a].

**Cystatin C**

Originally cystatin C was termed as γ-trace or post-γ-globulin isolated from human cerebrospinal fluid because of its basic nature and γ electrophoretic mobility [Barrett et al., 1984; Brzin, 1984]. It was also found in the urine in renal failure patients [Butler and Flynn, 1961] and ascetic and pleural fluids [Hochwald and Thornbecke, 1962]. Cystatin C was also detected in saliva, normal serum [Cejka and Fleischmann, 1973] and seminal plasma [Colle et al., 1976]. Preferentially abundant in cerebrospinal fluid, seminal plasma, milk, synovial fluid, urine, and blood plasma [Abrahamson et al., 1986] cystatin C has also been detected intracellularly in brain cortical nerves [Lofberg et al., 1981], normal and neoplastic neuroendocrine cells in the adrenal medulla [Lofberg et al., 1982], thyroid [Lofberg et al., 1983] and pituitary [Lofberg et al., 1983; Moller et al., 1985].

**Cystatin D**

Cystatin D a member of human cystatin multigene family and was cloned from a genomic library using cystatin C cDNA probe [Freije et al., 1991]. The inhibitor consists of 122 amino acids residues (Mr 13,885). The deduced amino acid composition includes a putative signal peptide and has 51-55% homology with either cystatin C or secretory gland cystatins S, SA and SN. It is a relatively neutral protein with pI in the range of 6.8- 7.0 [Freije et al., 1991]. It is expressed in parotid glands, saliva and tears [Balbin et al., 1994]. This tissue restricted expression is in marked contrast with a wider distribution of all other family 2 cystatins.

**Cystatin S**

Human saliva contains several low MW acidic proteins which include CPIs [Isemura et al., 1984b]. The first salivary inhibitor purified and sequenced was SAP-I (salivary acid protein) by Isemura et al. [1984a], which was renamed as ‘cystatin S’. It contains no phosphate, in contrast to other salivary proteins. This inhibitor has also been isolated from human submaxillary, submandibular and sublingual glands and found to be present in the serous cells of the parotid and submaxillary glands [Isemura et al.,]
1984b]. The protein has also been found in tears, serum, urine, bile, pancreas and bronchus [Isemura et al., 1986].

**Variants of cystatin S**

Several molecular variants of cystatin S have been studied by Isemura et al [1986] which differ in their N-terminal sequence and pI values. Differences in pI values resulted from phosphorylation of residues Ser3 and Ser1 in salivary cystatin [Isemura et al., 1991]. **Cystatin SN:** Originally known as cystatin SV or SA-1 [Abrahamson et al., 1986]. The protein consists of 121 amino acid residues (Mr 14,316). The pI values are in the range of 6.6-6.8. **Cystatin SA** consists of 122 amino acid residues (Mr 14,351) having acidic pI value of 4-6 [Isemura et al., 1991]. Cystatin SA isolated from saliva had N-terminal residue Glu [Isemura et al., 1986].

**Cystatin E**

Human cystatin E from amniotic fluid and fetal skin epithelial cell was identified and recombinant cystatin E isolated [Ni et al., 1997]. Human cystatin M is expressed by normal mammary cells and a variety of human tissues [Sotirpoulou et al., 1997]. Both proteins are identical and were renamed as cystatin E/M (MEROPS, the peptidase database). Recently, the expression of cystatin M/E was found to be restricted to the epidermis [Cheng et al., 2006] and is most probably identical to cystatin E/M.

**Cystatin F**

Cystatin F (leukocystatin) (MW 14,543) is primarily found in peripheral blood cells, T cells, spleen, dendritic cells and selectively, in hematopoietic cells [Ni et al., 1998; Halfon et al., 1998]. Cystatin F has an additional disulphide bridge, thus stabilizing the N-terminal part of the molecule [Ni et al., 1998]. It is the only cystatin synthesized and secreted as an inactive disulphide-linked dimeric precursor which becomes active following reduction to monomeric form [Schuttelkopf et al., 2006].

**TYPE 3 CYSTATINS (KININOGENS)**

Kininogens, the precursors of kinin, are large multifunctional glycoproteins in mammalian plasma and other secretions. Three different types of kininogens have been identified: high molecular weight kininogen (HK), low molecular weight
kininogen (LK) and T-kininogen an acute phase protein found only in rats [DeLa Cadena and Colman, 1991; Muller-Esterl, 1987]. Human HK and LK are single-chain proteins each composed of an N-terminal heavy chain, the kinin segment and a C-terminal light chain. The light and heavy chains are interconnected by disulphide bridges. The heavy chains and kinin segments of both kininogens have identical amino acid sequences while the light chains are different [DeLa Cadena and Colman, 1991; Salvesen et al., 1986a]. The heavy chain is composed of three cystatin domains [Fig. 1], D1-D3 [Salvesen et al., 1986a] with only D2 and D3 possessing papain inhibitory and D2 possessing calpain inhibitory activities. An inhibitory fragment, identical to the third domain of human kininogen, was isolated from human placenta and is inactivated by the lysosomal aspartic proteinase cathepsin D. Similarly HCC was also inactivated, suggesting a role for cathepsin D in regulating cysteine cathepsin activity [Lenarcic et al., 1991]. Both inhibitory domains of LK and HK are grouped in subfamily I25B of the cystatin superfamily [Rawlings et al., 2004a].

Other type 2 cystatins

There are a number of other cystatins or cystatin related proteins, which are structurally related to cystatins with no inhibitory activity against papain like enzymes [Turk and Turk, 2008]. CRES (Cystatin Related Epididymal Spermatogenic) protein [Sutton et al., 1999], testatin (expression restricted to mouse pre-Sertoli cells) [Tohonen et al., 1998], cystatin SC and cystatin TE-1 (expressed in testis and epididymis, respectively) [Li et al., 2002], and several other genes were found expressed specifically in the male reproductive tract [Hamil et al., 2002; Xiang et al., 2005; Shoemaker et al., 2000], indicating the existence of a new subgroup in the type two cystatins [Cornwall et al., 2003; Sutton-Walsh et al., 2006]. These CREStatins show homology to cystatins, with the exception of the two hairpin loops responsible for the cysteine proteinase inhibition. Their role could be regulation of proteolysis in the reproductive tract as well as protection against invading pathogens, as shown by cystatin 11 [Hamil et al., 2002]. The CRES protein tend to form oligomers [Horsten et al., 2007], similar to cystatin C [Janowski et al., 2001; Wahlbom et al., 2007] and stefin B [Zerovnik et al., 2002a; Jenko-Kokalj et al., 2007]. Another type 2 cystatin, cystatin 10, expressed in cartilage, localized in prehypertrophic and hypertrophic chondrocytes is known to be an inducer of chondrocyte maturation followed by apoptosis [Koshizuka et al., 2003]. A novel cystatin type 2 protein namely CLM
expressed widely in normal tissue playing role in hematopoietic differentiation or inflammation, different from CRES was characterized by Sun and coworkers [2003].

NEW MEMBERS OF THE CYSTATIN SUPERFAMILY

The feutins and histidine-rich glycoproteins (HRG) comprise fourth family of cystatins. The feutin family consists of two tandem cystatin domains. Bovine feutin was first characterized by Pedersen in 1944, and its relation to cystatin superfamily described in 1988 [Elzanowski et al., 1988]. Human feutin (α2-HS glycoprotein) was confirmed in 1987 [Dziegielewska et al., 1987; 1990; Dziegielewska and Brown, 1995]. Since then, protein and/or cDNA sequences have been reported for human, cow, pig, rat, mouse, Habu snake, feutins [Brown and Dziegielewska, 1997]. Almost all the feutin sequences contain 12 cysteine residues, showing homology to the cystatins and cystatin domains in kininogens [Dziegielewska and Brown, 1995]. HRG has been characterized in the plasma of man, mouse, rabbit, cow and pig [eung, 1993], sharing good sequence homology with human and bovine HMW kininogen [Koide et al., 1986]. A large number of proteins have been discovered recently, which possess cystatin domains e.g. latexin [Aagaard et al., 2005]. However, feutin, HRG and latexin all seem to lack CPI activity.

Thyropins constitute a new family of papain-like CP inhibitors [Lenarcic and Bevec, 1998], classified as family I31 [Rawlings et al., 2004a]. The p41 invariant chain (Ii)-fragment of the MHC class II-II complex 104, 105 and equistatin from the sea anemone [Lenarcic et al., 1997] are best characterized members of this family. Thyropins show inhibitory activity against CPs and also towards aspartic and metalloproteinases [Mihelic and Turk, 2007; Lenarcic and Turk, 1999]. Tick cystatins: Syalostatin L [Kotsyfakis et al., 2006] and syalostatin L2 [Kotsyfakis et al., 2007] have been characterized from salivary glands of the tick Ixodes scapularis. Both show 75% sequence identity and inhibit cathepsin L with a Ki of 4.7 nM and cathepsin V with Ki of 57 nM. Staphostatins are specific inhibitors of staphylococcal CPs. Three members of this family have been described-staphostatins A and B from Staphylococcus aureus and staphostatin A from Staphylococcus epidermidis [Filipek et al., 2003]. Clitocybin is a new type of CPI from a mushroom appearing to be related to fungal lectins and hence a new family of CPIs is suggested for them called mycocypins [Brzin et al., 2000]. Chagasin is a cysteine proteinase inhibitor from
Trypanozoma cruzi inhibiting both cruzipain and papain, but has no homology with cystatins [Monteiro et al., 2001].

**Phytocystatins:** In plants, inhibitors of CPs are known as phytocystatins. They contain the QXVXG region of type 2 cystatins, but also resemble stefins in the absence of disulphide bonds [Arai et al., 2002], providing a transitional link between type 1 and type 2 cystatins. There are numerous phytocystatins expressed and characterized on the protein level from corn [Abe et al., 1992], rice [Chen et al., 1992], soyabean [Lalitha et al., 2005], sugarcane [Oliva et al., 2004] and others. C-terminal extended phytocystatins were found as bifunctional inhibitors of papain and legumain [Martinez et al., 2007]. In addition, a “multicystatin” containing two cystatin like domains were isolated from cowpea leaves [Diop et al., 2004], tomato leaves [Wu and Haard, 2000]. Also there are certain plant proteins like monellin which lack the CPI activity but have a cystatin like three dimensional structure [Grzonka et al., 2001]. Phytocystatins and other inhibitors are important for plant defence response to insect predation, may act to resist infection by some nematodes [Koiwa et al., 1997], play a crucial role in response to various conditions [Diop et al., 2004; Brzin and Kidric, 1995] and show great potential tools for genetically engineered resistance of crop plants against pests [Aguiar et al., 2006].

**Variant cystatins**

Divergent cystatins showing significant homology to stefins, cystatins and kininogens have been expressed/purified and characterized from venom of African puff adder (*Bitis arietans*) [Evans and Barrett, 1987]; from perilymph of flesh fly larvae [Suzuki and Natori, 1985]; from *Drosophila melanogaster* [Delbridge and Kelly, 1990]. Some of the mammalian and non mammalian sources from where CPIs have been isolated are summarized in Table 1.

**1.5 EVOLUTION**

The first two proposed evolutionary dendrograms for CPIs were made based on a small number of members of the cystatin superfamily [Muller-Esterl et al., 1985b; Salvesen et al., 1986b]. The new proposed evolutionary dendrograms followed the evolution of the proteins of the cystatin superfamily along four lineages, with special
TABLE 1: CPIs FROM SOME MAMMALIAN AND NON-MAMMALIAN SOURCES

<table>
<thead>
<tr>
<th>Source</th>
<th>Tissue</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beef</td>
<td>Spleen</td>
<td>Brzin et al., 1982</td>
</tr>
<tr>
<td>Bovine</td>
<td>Brain</td>
<td>Aghajanyan et al., 1988</td>
</tr>
<tr>
<td></td>
<td>Hoof</td>
<td>Tsushima et al., 1996</td>
</tr>
<tr>
<td></td>
<td>Colostrums</td>
<td>Hirado et al., 1985</td>
</tr>
<tr>
<td>Dog</td>
<td>Colostrum</td>
<td>Poulik et al., 1981</td>
</tr>
<tr>
<td>Horse show crab</td>
<td>Parotid gland &amp; Kidney</td>
<td>Tsushima et al., 1996</td>
</tr>
<tr>
<td></td>
<td>Hemocytes</td>
<td>Sekine and Poulik, 1982</td>
</tr>
<tr>
<td>Human</td>
<td>Liver</td>
<td>Green et al., 1984</td>
</tr>
<tr>
<td></td>
<td>Spleen</td>
<td>Jarvinen and Rinnie, 1982</td>
</tr>
<tr>
<td></td>
<td>Placenta</td>
<td>Rashid et al., 2006a</td>
</tr>
<tr>
<td>Rabbit</td>
<td>Liver</td>
<td>Pontremoli et al., 1983</td>
</tr>
<tr>
<td></td>
<td>Skin</td>
<td></td>
</tr>
<tr>
<td>Rat</td>
<td>Brain</td>
<td>Kopitar et al., 1983</td>
</tr>
<tr>
<td><em>Ixodes scapularis</em></td>
<td>Salivary gland</td>
<td>Kotsyfakis et al., 2006 &amp; 2007</td>
</tr>
<tr>
<td><em>Staphylococcus aureus</em></td>
<td>-</td>
<td>Filipek et al., 2003</td>
</tr>
<tr>
<td>&amp; <em>epidermidis</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Trypanosoma cruzi</em></td>
<td>-</td>
<td>Monteiro et al., 2001</td>
</tr>
<tr>
<td><em>Fasciola hepatica</em></td>
<td>-</td>
<td>Khaznadji et al., 2005</td>
</tr>
<tr>
<td>Goat</td>
<td>Kidney</td>
<td>Zehra et al., 2005</td>
</tr>
<tr>
<td></td>
<td>Brain</td>
<td>Sumbul &amp; Bano, 2006</td>
</tr>
<tr>
<td></td>
<td>Lung</td>
<td>Khan &amp; Bano, 2009a</td>
</tr>
<tr>
<td><em>Spirometra erinacei</em></td>
<td>-</td>
<td>Priyadarshini &amp; Bano, 2009</td>
</tr>
<tr>
<td>Yellow croaker</td>
<td>Spleen</td>
<td>Chung &amp; Yang, 2008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Li et al., 2009</td>
</tr>
</tbody>
</table>
attention that duplication of cystatin like segments has played important contribution
to the understanding of the evolution of cystatins. According to the scheme of Muller-
Esterl et al. [1985b] constructed on the basis of sequence homology, the diversity of
CPIs has evolved from two ancestral building blocks ‘A’ and ‘B’. The stefin
progenitor represents the whole superfamily comprising a single ‘A’ unit. Cystatin
acquired a second element B, possibly by gene fusion, thus forming ‘AB’ unit. Gene
triplcation of the archetype inhibitor generated the kininogen heavy chain which
contains 3 cystatin like copies (AB)₃. The proposed evolutionary pathway also
contained a ‘missing link’, a two cystatin domain protein that evolved from the
cystatins by duplication, with two candidates for such a protein: feutin and HRG. This
scheme however seem unlikely most importantly because neither domain in
feutins/HRG is inhibitory but two domains of kininogens have inhibitory activity. If
feutin/HRG were the ‘missing link’, then the kininogens which have evolved from the
two domain protein would have to re-evolve their proteinase inhibitory activity and
proposed the following scheme for cystatin superfamily evolution, with features
similar to Muller-Esterl et al., [1985b] scheme but with a new missing link, a two
cystatin domain protein in which both the domains were functional cysteine
proteinase inhibitors. From it, the kininogens, feutins, and HRG could have evolved
separately or perhaps in parallel and retained or lost their proteinase-inhibitory
activity and active site sequences. This scheme [Fig. 3] draws support from the
observation of conserved sequences immediately around the cysteine at the C-
terminus of the feutins, HMW-kininogens and HRG, again suggesting a common
origin for these three proteins. Based on this Lee et al., [2009] have recently grouped
feutins, HRG and kininogens in a single family, type 3 cystatins.

1.6 STRUCTURE OF CYSTATINS

PRIMARY STRUCTURE

Most of the members of cystatin superfamily are polypeptides of 98-126 amino acid
residues with Mr values in the range of 11-14 kDa. As regards to the amino acid
composition of cystatins few distinctive features can be attributed to the subfamilies.
Stefins are devoid of disulphide linkages (human cystatin A and rat cystatin α lack
cysteine residues while human cystatin B and rat cystatin β have 1 and 2 cysteine
Fig. 3  Evolution of cystatin superfamily

A. Scheme from Muller-Esterl et al. [1985b]
B. Scheme proposed by Brown & Dziegielewksa [1997]
(A) Stefin

Cystatin

(AB)_3

Kininogen

(B)

archetypal 3-exon animal cystatin

missing link

fetuin

HRG

HMW kininogen

no cysteine protease inhibitor

missing link

cystatin

cysteine protease inhibitor

missing link
residues, respectively) and tryptophan. Turk et al. [1993] however reported the presence of tryptophan in stefin C. A unique feature of stefin B is the conserved QVVGAG region in the stefins of mammalian origin, with Val54 replaced by Leu54. Ni et al. [1998] reported the presence of an additional disulphide bridge in cystatin F, for stabilizing the N-terminal part of the molecule in addition to the presence of second tryptophan residue, along with the conserved Trp 106, characteristic of type 2 cystatins. The alignment of sequences of cystatins reveals common features, significant to the structure and activity of the proteins. Four residues are common to all the sequences of cystatins and inhibitory kininogen segments: Gly9, Gln53, Val55 and Gly57. These residues are considered to be of functional importance since they are absent from the non-inhibitory segment D1 of kininogens. Another six conserved residues are Val47, Val55, Ala56, Tyr60, Cys71 and Tyr100. The segment Gln53 to Gly57 is the most highly conserved region.

SECONDARY STRUCTURE

The crystalline form of chicken cystatin reported by Bode et al. [1988] revealed a new fold, the cystatin fold which is, a five stranded anti-parallel β-sheet wrapped around the central N-terminal helix [Fig. 2]. This fold has been shown to exist in HCC, chicken cystatin, cystatin D, as well as in family 1 cystatins A and B [Martin et al., 1994; 1995; Alvarez-Fernandez et al., 2005; Stubbs et al., 1990]. An appending segment of partial α-helical geometry is present in chicken cystatin [Saxena and Tayyab, 1997], but absent in HCC [Bode et al., 1988]. Tryptophan was found only in the second hairpin loop of cystatins [Bode et al., 1988]. A unique feature was observed in crystal structure of cystatin F in its dimeric ‘off’ state. The two monomers interacted in a fashion not seen before for cystatins or cystatin like proteins, crucially dependent on an unusual intermolecular disulphide bridge. The core sugars for one of the two N-linked glycosylation sites for cystatin F are well ordered and probably their conformation and interactions with the protein modulate its inhibitory properties in particular its reduced affinity toward asparaginyl endopeptidase compared with other cystatins [Schuttelkopf et al., 2006].

There is considerable similarity between the structural features of stefins A and B [Fig. 4], but there are also some important differences in the regions which are fundamental to proteinase binding. The difference primarily consists of the two regions of high conformational heterogeneity in free stefin A which correspond in
Fig. 4  (A) Three dimensional structure of stefin A (where, A, B, C, D, E are the five antiparallel β sheets strands)
(B) Three dimensional structure of stefin B (antiparallel β sheets are in green colour and α-helix is in red colour)
stefin B to two of the components of the tripartite wedge that docks into the active site of target proteinases. These regions which are mobile in solutions are the five N-terminal residues and the second binding loop. In the bound conformation of stefin B they form a turn and a short helix, respectively.

Circular dichroism and computer prediction of secondary structure from the sequence indicates that chicken cystatin has about 20% α-helix, 42% β-structure, 24% β-turn and 12% random coil [Schawbe et al., 1984]. Recombinant human cystatin A in good comparison to cystatin A, in far UV-CD spectrum revealed ~45% β-structure and a low α-helix content (~15%) [Pol et al., 1995].

1.7 INHIBITION OF PROTEINASES

Specificity

Cystatins are highly specific for CPs except for thyropins which show inhibitory activity against aspartic and metalloproteinases [Mihelic and Turk, 2007; Lenarcic and Turk, 1999]. However there are few cystatins capable of inhibiting mammalian legumain [Alvarez-Fernandez et al., 1999] and calpains [Crawford, 1987]. To date, none of the cytoplasmic inhibitors have been tested on ubiquitin processing and recycling proteinases [Keppler, 2006]. Stefin A and B are potent inhibitors of papain, cathepsin L, S and H but have decreased activity against cathepsin B [Musil et al., 1991]. Type 2 cystatins are important endogenous inhibitors of papain like CPs including cathepsins, parasite proteinases like cruzipain and mammalian legumain [Turk et al., 2005; Turk and Bode, 1991]. HCC and chicken cystatin inhibit papain, cathepsin L and S [Abrahamson et al., 2003]. HCC shows strong inhibitory capacity for rapid binding thus neutralizing proteinase activity in an emergency inhibition [Turk et al., 2005]. It also inhibits cruzipain, suggesting its possible defensive role after infection [Stoka et al., 1995]. Cystatin F inhibits cathepsin F, K, V, S, L and H [Langerholc et al., 2005] and weakly legumain [Alvarez-Fernandez et al., 1999]. More recently it was found that the intracellular form of cystatin F, after N-terminal truncation of the first 15 residues including cysteine, inhibits cathepsin C [Hamilton et al., 2008]. Cystatin D inhibits cathepsin S, H and L but not cathepsin B or pig legumain [Alvarez-Fernandez et al., 2005]. Human cystatin E/M inhibits papain, cathepsin B, L, V and legumain [Ni et al., 1997; Sotiropoulou et al., 1997; Cheng et al., 2006; Alvarez-Fernandez et al., 1999]. Clostripain (proteinase not belonging to
papain family) is also inhibited by cystatins [Barrett et al., 1986].

**Kinetic behaviour**

Cystatins are the first group of protein inhibitors of CPs for which the mechanism of inhibition was investigated. All the cystatins are non-covalent, competitive, reversible, tight binding inhibitors which inhibit the target enzymes in micromolar to picomolar range [Turk et al., 1997]. They form tight equimolar complexes with CPs [Anastasi et al., 1983]. Some of the reported values of equilibrium constants for dissociation of complexes between human cystatins and lysosomal CPs are summarized in Table 2. The affinity differences can be explained by the differences in the active site regions of endo- and exopeptidases. The access of the inhibitor to the active site of exopeptidases is partially obstructed by occluding loops in cathepsin B [Musil et al., 1991], cathepsin X [Guncar et al., 2000], propeptide parts in cathepsin H [Guncar et al., 1998] and cathepsin C [Turk et al., 2001].

**Reactive site and mechanism of action**

It has been established that no disulphide bond is formed between the active site cysteine residue and the inhibitor because the complexes dissociated when denatured without reduction as was found in chicken cystatin [Nicklin and Barrett, 1984] and kininogens [Gounaris et al., 1984]. The complex formation is accompanied by pronounced spectroscopic changes [Bjork et al., 1989]. On the basis of cystatin domain structure, it was proposed that there are three regions crucial for interaction with proteinases: the amino terminus and two β-hairpin loops, one in the middle and one in the C-terminal segment of the protein. The first loop contains a QXVXG sequence conserved in almost all inhibitory members of cystatins, whereas the second loop contains a P-W motif, which is also highly conserved [Table 3]. Both these loops and the amino terminus forms a wedge shaped edge, which is highly complementary to the active site of the enzyme. The N-terminally truncated forms of chicken cystatin confirmed the crucial importance for the binding of the residues preceding the conserved Gly-9 residue [Machleidt et al., 1989]. The essential interactive elements of this hypothetical complex are shown in figure 5. Complex formed on interaction of stefin B with cathepsin H is shown in figure 6.
TABLE 2: EQUILIBRIUM CONSTANTS FOR DISSOCIATION (Ki) OF COMPLEXES BETWEEN HUMAN CYSTATINS AND CHICKEN CYSTATIN WITH LYOSOMAL CYSTEINE PROTEINASES (PAPAIN, HUMAN CATHEPSINS AND CRUZIPIAIN)

<table>
<thead>
<tr>
<th>CPI</th>
<th>Papain</th>
<th>Cathepsin B</th>
<th>Cathepsin H</th>
<th>Cathepsin L</th>
<th>Cruzipain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cystatin A</td>
<td>0.019</td>
<td>8.2</td>
<td>0.31</td>
<td>1.3</td>
<td>0.0072</td>
</tr>
<tr>
<td>Stefin B</td>
<td>0.12</td>
<td>73</td>
<td>0.58</td>
<td>0.23</td>
<td>0.060</td>
</tr>
<tr>
<td>Cystatin C</td>
<td>0.00001</td>
<td>0.27</td>
<td>0.28</td>
<td>&lt;0.005</td>
<td>0.014</td>
</tr>
<tr>
<td>Cystatin D</td>
<td>1.2</td>
<td>&gt;1000</td>
<td>7.5</td>
<td>18</td>
<td>n.d.</td>
</tr>
<tr>
<td>Cystatin E/M</td>
<td>0.39</td>
<td>32</td>
<td>n.d.</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>Cystatin F</td>
<td>1.1</td>
<td>&gt;1000</td>
<td>n.d.</td>
<td>0.31</td>
<td>n.d.</td>
</tr>
<tr>
<td>Cystatin SA</td>
<td>0.32</td>
<td>n.d.</td>
<td>n.d.</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>Cystatin SN</td>
<td>0.016</td>
<td>19</td>
<td>n.d.</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>Chicken cystatin</td>
<td>0.005</td>
<td>1.7</td>
<td>0.06</td>
<td>0.019</td>
<td>0.001</td>
</tr>
<tr>
<td>L-kininogen</td>
<td>0.015</td>
<td>600</td>
<td>0.72</td>
<td>0.017</td>
<td>0.041</td>
</tr>
<tr>
<td>H-kininogen</td>
<td>0.02</td>
<td>400</td>
<td>1.1</td>
<td>0.109</td>
<td>n.d.</td>
</tr>
</tbody>
</table>

n.d. (not determined). Ki values for human cystatins [Abrahamson et al., 2003], chicken cystatin [Barrett et al., 1986] and cruzipain inhibition by cystatins [Stoka et al., 1995].
Table 3: **CONSERVED AMINO-ACID RESIDUES IN BINDING SEGMENTS OF HUMAN CYSTATINS**

<table>
<thead>
<tr>
<th>Cystatin</th>
<th>N-terminus</th>
<th>I loop</th>
<th>II loop</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>MIPGG</td>
<td>QVVAG</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>AcMMCGA</td>
<td>QVVAG</td>
<td>VPWQ</td>
</tr>
<tr>
<td>C</td>
<td>RLVGG</td>
<td>QIVAG</td>
<td>VPWE</td>
</tr>
<tr>
<td>D</td>
<td>TLAGG</td>
<td>QIVAG</td>
<td>VPWQ</td>
</tr>
<tr>
<td>E</td>
<td>RMVGE</td>
<td>QLVAG</td>
<td>VPWE</td>
</tr>
<tr>
<td>F</td>
<td>VKPGF</td>
<td>QIVKG</td>
<td>VPWL</td>
</tr>
<tr>
<td>S</td>
<td>IIPGG</td>
<td>QTFGG</td>
<td>VPWE</td>
</tr>
<tr>
<td>SA</td>
<td>IIEGG</td>
<td>QIVGG</td>
<td>VPWE</td>
</tr>
<tr>
<td>SN</td>
<td>IIPGG</td>
<td>QTVGG</td>
<td>VPWE</td>
</tr>
<tr>
<td>H-kininogen</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-domain</td>
<td>(QESQS)</td>
<td>(TVGSD)</td>
<td>(RSST)</td>
</tr>
<tr>
<td>2-domain</td>
<td>DCLGC</td>
<td>QVVAG</td>
<td>(DIQL)</td>
</tr>
<tr>
<td>3-domain</td>
<td>ICVGC</td>
<td>QVVAG</td>
<td>VPWE</td>
</tr>
<tr>
<td>L-kininogen</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-domain</td>
<td>(QESQS)</td>
<td>(TVGSD)</td>
<td>(RSST)</td>
</tr>
<tr>
<td>2-domain</td>
<td>DCLGC</td>
<td>QVVAG</td>
<td>(DIQL)</td>
</tr>
<tr>
<td>3-domain</td>
<td>ICVGC</td>
<td>QVVAG</td>
<td>VPWE</td>
</tr>
</tbody>
</table>

* Sequences in parenthesis correspond to the appropriate binding sequences of cystatins.

b Grzonka et al., 2001.
Fig. 5  Scheme of the proposed model for the interaction of chicken egg white-cystatin and papain

(Adapted from Turk and Bode, FEBS Lett 1991; 285:213-219)
Fig. 6  Three dimensional structure of the complex formed between stefin A and cathepsin H

Binding of stefin A into cathepsin H active site. Stefin A fold is shown as a green chain trace. Whereas cathepsin H fold is shown in yellow. Cathepsin H mini-chain residues are shown as red sticks which are thicker for the main chain. The mini-chain is attached to the body of cathepsin H with a disulfide shown as red yellow chain. The identified carbohydrate rings are shown in cyan. The N-terminus of stefin A displaces the C-terminus of the minichain by pushing its residues outside the binding cleft. Adopted from Turk and Turk, Acta Chim Slov 2008; 55:727-738.
Bode et al. [1988] demonstrated that the major contribution is from the first hairpin loop containing QVVAG sequence [Turk et al., 1985]. According to the model the N-terminal segment of cystatin which is more flexible bridges over the active site Cys 25 of papain without completely burying it and additionally the side chain of Leu 8 binds to S2 subsite of papain which determines the substrate specificity of papain [Asboth et al., 1988]. This was supported by Brzin et al. [1984] who demonstrated that the truncated form of HCC starting with Leu-Val before Gly-11 (corresponding to Gly-9 of chicken cystatin) has virtually the same affinity for papain as the full length form whereas the truncated form starting with Gly-12 has been reported to show 1000 fold weaker inhibition [Abrahamson et al., 1987a]. However, Nycander and Bjork [1990] emphasized the role of Trp-104 in the inhibition of CP. According to their model, Trp-104 of cystatin interacts primarily with two Trp side chains in the active side cleft of papain, Trp 177 and Trp 181, in such a manner that the indole ring of Trp-104 stacks on the side chain of Trp 177 and the edge lies on the indole ring of Trp 181. A two step mechanism of inhibition of the lysosomal CP cathepsin B by its endogenous inhibitor cystatin C was observed by Nycander et al. [1998]. An initial weak interaction in which N-terminal of the inhibitor binds to the proteinase is followed by a conformational change. Subsequently, the occluding loop of the proteinase that partially obscures the active site is displaced by the inhibitor bringing about another conformational change. The presence of occluding loop of cathepsin B renders it much less susceptible to inhibition by cystatin than other proteinases. A similar two step binding of cystatin A to CP was suggested by Estrada and Bjork [2000].

The flexible N-terminal region of the cystatin binds independently to the target proteinases after the binding of hairpin loops. It is interesting to note that the replacement of the three N-terminal residues preceding the conserved Gly of stefin A by the corresponding 10-residues long segment of cystatin C increased the affinity of the inhibitor for cathepsin B by about 15-fold [Pavlova and Bjork, 2003], suggesting that the inhibitory potency of cystatin can be substantially improved by protein engineering. The crystal structure of human stefin A-porcine cathepsin H complex showed small distortion of the structure upon formation of the complex [Jenko et al., 2003]. In addition to the structurally derived data, the contribution of the individual residues within proteinase binding region of cystatins was additionally investigated by mutational analysis and kinetic studies performed by several different groups [Pol and Bjork, 2001; Auerswald et al., 1994; Estrada et al., 1999; Pavlova et al., 2000].
1.8 BIOLOGICAL ASPECTS AND PATHOPHYSIOLOGY OF CYSTATINS

Proteinases and their natural inhibitors may co-exist at different levels of cellular evolution. Disturbing the harmony of the normal balance of enzymatic activities of proteinases and their natural inhibitors may lead to severe biological effects. Cystatins constitute a powerful regulatory system for endogenous CPs which are often secreted or leaked from the lysosomes of dying and diseased cells [Ekiel et al., 1997]. Besides regulation of the enormous hydrolytic potential of CPs, plethora of actions has now been ascribed to cystatins. They are known to play important roles in various pathophysiological conditions such as sepsis [Assfalg-Machleidt et al., 1988], cancer [Cox, 2009], rheumatoid arthritis [Trabandt et al., 1991], purulent bronchiectasis [Buttle et al., 1990], multiple sclerosis [Bever and Garver, 1995], muscular dystrophy [Sohar et al., 1988], etc. which indicate that a tight enzyme regulation by cystatin is a necessity in the normal state.

Cystatins and cancer: Cathepsins involved in the degradation of extracellular matrix facilitate the growth, invasion and metastasis of tumour cells and also in tumour angiogenesis [Gocheva and Joyce, 2007; Mohamed and Sloane, 2006; Turk et al., 2004; Vasiljeva et al., 2006]. A broad spectrum of cysteine proteinase inhibitor was shown to inhibit tumour angiogenesis [Joyce et al., 2004]. Generally, cathepsin to cystatin ratio is found to be increased in most tumour types compared to normal tissues [Paraoan et al., 2009; Rivenbark and Coleman, 2009]. Elevated TPI level in various tumour types have been correlated to better prognosis like, stefin A positive breast cancer patients are less likely to develop distant metastasis [Parker et al., 2008], stefin A and B in non small cell lung cancer [Werle et al., 2006], cystatin SN is upregulated in gastric cancer [Choi et al., 2009], cystatin C [Sokol and Schiemann, 2004], cystatin M [Zhang et al., 2004], cystatin F [Utosunomiya et al., 2002], were found to be expressed in epithelial and mesenchymal tumour cells. Cystatin M is often hailed as tumour suppressor.

Stefin A and cystatin C overexpression has been shown to inhibit cancer cell invasion and metastasis [Li et al., 2005; Kopitz et al., 2005]. Cystatins may also inhibit cell migration by interfering with cell signalling pathways, by direct cathepsin and calpain inhibition [Cox, 2009]. Cathepsins B, L and S promote tumour growth in a
murine model of pancreatic tumourigenesis [Gocheva et al., 2006]. The tumour microvascular density declined by about half in pancreatic tumours in cathepsin B or cathepsin S null mice. Significant increases in apoptosis were also noted in cathepsin B, L and S null pancreatic tumours.

**Cystatins and neurodegeneration:** Though CPs are implicated in various pathologies of brain, there are only few studies concerning the role of cystatins in pathologies of brain. Only two genetic diseases are known in which mutations in cystatin C and stefin B are associated with disease status, **Hereditary cystatin C amyloid angiopathy (HCCAA)** the first human disorder known to be caused by deposition of cystatin C amyloid fibrils in walls of brain arteries leading to single to multiple strokes with fatal outcomes [Jensson et al., 1987]. The amyloid deposited is composed mainly of the Leu 68 Gin variant of cystatin C [Wei et al., 1998] and is associated with mutation in cystatin C gene [Palsdottir et al., 2006]. Normal cystatin C may protect pathogenesis of Alzheimer’s disease by binding to soluble amyloid-β-peptide and preventing its deposition [Mi et al., 2007].

**Progressive myoclonus epilepsy [EPM1]** is exhibited by a group of inherited diseases characterized by myoclonic seizures, generalized epilepsy and progressive neurological degeneration caused by mutations in cystatin B gene (in the conserved QVVAG region) [Pennacchio et al., 1997; Joensuu et al., 2007].

**Cystatins and cell death:** Cystatins are shown to be involved in normal cell apoptosis, and most dramatically in selective tissue type for example in EPM1 [Lieuallen et al., 2001]. In fibrosarcoma cells, cystatins regulate cell death in response to TNFα [Foghsgaard et al., 2001]. An elevated cystatin C/cathepsin B ratio was found to be associated with chemoresistance in non-small cell lung carcinoma patients [Petty et al., 2006]. Intracellular cystatins normally inhibit low level lysosomal leakage. High cystatin levels are expected to be protective for general cathepsin mediated cell death.

**Cystatins and immunomodulation:** Cystatins have emerged as effector molecules of immunomodulation [Zavasnik-Bergant, 2008]. They can stimulate nitric oxide release from macrophages [Verdot et al., 1999]; modulate respiratory burst and phagocytosis in neutrophils [Leung-Tack et al., 1990]; and interleukin, cytokine
production in T-cells and fibroblasts [Schierack et al., 2003; Kato et al., 2002; 2004]. Most of these functions operate via putative cell surface cystatin-binding molecules or membrane domains [Kato et al., 2002]. Cystatin C has been shown to be a TGFβ receptor antagonist and TGFβ signalling pathway blocker [Sokol and Schiemann, 2004; Sokol et al., 2005]. Type 2 cystatins are also known to increase interleukin-6 expression in fibroblasts and splenocytes [Kato et al., 2000]. Cystatin C is a potent, reversible inhibitor in vitro of the human lysosomal CPs e.g., cathepsin S (Ki = 8 pM), cathepsin L (Ki = 8 pM) and cathepsin H (Ki = 220 pM). These proteinases are located all along the endocytic pathway of dendritic cell and are involved in the controlled proteolysis associated with the degradation of antigenic peptides [Pluger et al., 2002]. Cystatin F by targeting cathepsin C is known to regulate diverse immune cell effector functions [Hamilton et al., 2008].

**Cystatins as antimicrobial and antiviral agents:** Horse-shoe crab hemocyte cystatin has antimicrobial activity against Gram negative bacteria with IC50s against *S. typhimurium, E. coli* and *K. pneumoniae* in the 80-100 µg/ml range [Agarwala et al., 1996]. Both chicken and human cystatins were found to inhibit the growth of *P. gingivalis* with an IC50 of 1.1 and 1.2 fM, respectively [Blank et al., 1996]. Cystatin C is also an effective inhibitor of replication of coronavirus [Collins and Grubb, 1998]. Sialostatin L displays anti-inflammatory role and inhibits proliferation of cytotoxic T-lymphocytes [Kotsyfakis et al., 2006].

**Cystatin C in clinical diagnostics:** Cystatin C was the first protein to be used in clinical diagnostics. Levels of cystatin C in various body fluids is used as a barometer of disease [Shah and Bano, 2009]. Recent studies indicate that it is a better marker of glomerular filtration rate and is a stronger predictor of cardiovascular disease and mortality than serum creatinine [Fried, 2009]. Cystatin C levels can also be used to reflect the characteristics of peritoneal membrane in dialysis patients [Al-Wakeel et al., 2009]. Korolenko et al. [2008] recently found that serum cystatin C concentration can be used as one of the prognostic criteria in patients with several kinds of hemoblastoses. IL-6 levels along with that of cystatin C may be regarded as markers of increased osteoblastic activity associated to bisphosphate treatments in prostrate cancer patients with bone metastases [Tumminello et al., 2009]. Yang et al. [2009] found that its levels decrease significantly in cerebrospinal fluids of patients with
Guillain-Barre syndrome and may be involved in its pathophysiology. Cystatin B was found to be specifically over expressed in most hepatocellular carcinomas and alone or in combination with α-fetoprotein may be a useful marker for diagnosis of the diseases [Lee et al., 2008].

Role of cystatins in other diseases: Cystatins are now known to participate in neuronal differentiation [Taupin et al., 2000]. Numerous studies have demonstrated that cathepsin K, L and S are involved in elastic fibre degradation, associated with the development of different pathological conditions of cardiovascular system. Elastinolytic activities of cathepsin K, L and S can be blocked by cystatins [Novinec et al., 2007]. Equistatin is known to inhibit the growth of the red flour beetle Tribolium castaneum [Oppert et al., 2003], suggesting to be promising candidate for the transgenic seed technology to enhance seed resistance to storage pests. Heparin binding and cell-binding domain 5 (light chain) of H-kininogen has antibacterial activity against E. coli, P. aeruginosa and Enterococcus faecalis [Andersson et al., 2005]. Bradykinin induces dendritic cell maturation and can modulate innate or adaptive immunity [Aliberti et al., 2003; Scharfstein et al., 2007]. Cystatin C appears to be up-regulated in response to injury in the brain [Shah and Bano, 2009]. Cystatin M/E is a key molecule in a biochemical pathway that controls skin barrier formation by the regulation of both crosslinking and desquamation of stratum corneum [Zeeuwen et al., 2009]. Lower cystatin C level is also implicated in retinal degeneration in (rd1) mouse model of retinitis pigmentatio [Ahuja et al., 2008]. An imbalance in cathepsin B/cystatin C level may contribute to the progression of pelvic inflammatory disease [Tsai et al., 2009]. There are many other diseases with decreased cystatin levels, such as inflammatory diseases, osteoporosis, arthritis, and diabetes as well as a number of other neurodegenerative diseases [Turk et al., 2008].

1.9 PANCREAS

Pancreas, the capital of proteolytic power, plays important role in digestion and in maintenance of glucose homeostasis in body. A soft lobulated, greyish-pink gland, 12-15 cm long (in humans), it extends nearly transversely across the posterior abdominal wall, behind stomach, from duodenum to spleen. The pancreas is composed of two separate types of glandular tissue [Fig. 7], main mass of which is exocrine, embedded in which are clusters of endocrine cells constituting pancreatic
Fig. 7  Anatomy of pancreas

islets. The exocrine part of pancreas is a lobulated, branched, acinar gland [Beck and Sinclair, 1971]. The acinar cells are zymogenic containing secretory granules laden with powerful enzymic constituents of pancreatic secretion. The endocrine pancreas consists of pancreatic islets (of Langerhans) spheroidal or ellipsoidal clusters or solitary or randomly embedded in the exocrine part of the pancreas [Heitz et al., 1976]. The islets contain three major types of cells, alpha [glucagon secreting; Baum et al., 1962], beta [most abundant and insulin secreting; Lacy and Davies, 1957] and delta [somatostatin secreting; Orci et al., 1975] cells. In addition, PP cell is also present in pancreas secreting hormone pancreatic polypeptide [Buffa et al., 1977].

Two major functions of the pancreas are release of insulin and glucagon in response to body glucose levels indirectly regulating protein, carbohydrate and lipid metabolism and acting as a switch between carbohydrate and lipid metabolism. Dysregulation of insulin and glucagon production causes diabetes mellitus, hyperinsulinism and hyperglycemia, respectively. Pancreatic acini secrete digestive enzymes and large volumes of sodium bicarbonate responsible for digestion of all three major food types- proteins, carbohydrates and fats. The proteolytic enzymes of the pancreatic juice are secreted in their zymogenic form.

1.10 PROTEIN UNFOLDING STUDIES

Proteins are synthesized as a linear chain of amino acids which in order to become biologically active must fold and adopt one out of an enormous number of possible conformations. This conformation referred to as native state, exists in solution as a very compact and highly ordered structure. Under physiological conditions the native (folded) and the denatured (unfolded) states of a protein are in equilibrium and the free energy change, $\Delta G$, for the equilibrium reaction

$$\text{Native (N) } \rightleftharpoons \text{Denatured (D)}$$

is referred to as the conformational stability of a protein. The determinants of native state stability in aqueous solutions are the amino acids sequence of protein as well as the variable conditions of pH, temperature, and the concentration of salts and ligands [Alber, 1989; Pace, 1990]. Although the native conformation is essential for activity, the conformational stability is remarkably low. The native state of most naturally occurring proteins is only about 5-15 kCal/mol more stable than its unfolded conformations [Pace, 1975]. The conformational stability of proteins (monomeric or
multimeric) can be measured by equilibrium unfolding studies using guanidine hydrochloride (GdnHCl) or urea [Pace, 1986], the two agents commonly employed as protein denaturants. Analysis of the solvent denaturant curves using these denaturants can provide measure of the conformational stability of protein [Pace, 1986; Yao and Bolen, 1995]. Protein unfolding/folding studies in GdnHCl and urea solutions have focused on the identifications of equilibrium and kinetic intermediates [Kim and Baldwin, 1990; Jaenicke, 1987]. Equilibrium denaturation studies using physical/chemical denaturants have been very useful in understanding the structure, stabilization and folding of small monomeric proteins [Tanford, 1968; Pace, 1986; 1990]. Lately, these techniques have been applied to oligomeric proteins also [Neet and Timm, 1994; Prakash et al., 2002; Akhtar et al., 2002]. Though inter- and intrasubunit reactions in oligomeric and monomeric proteins are of same physical nature, the denaturation/renaturation reactions in oligomeric proteins are more complex, usually multiphasic processes with stabilized partially folded intermediates [Jaenicke and Lilie, 2000; Hornby et al., 2000], occurring either by sequential or concerted mechanisms [Jaenicke, 1987; Seckler and Jaenicke, 1992].

With dimeric (or oligomeric) proteins additional modes of stabilization are available at the quaternary structural level. A general three-state model of the equilibrium dissociation and unfolding of the dimeric protein involving native dimer (N₂), native monomer (N) or a monomeric intermediate (I) and denatured monomer (D) is described by following equation

\[ N₂ \leftrightarrow 2N \leftrightarrow 2D \quad \text{or} \quad N₂ \leftrightarrow 2I \leftrightarrow 2D \]

Completely unfolded dimers (D₂) are not likely to exist, but evidence for partially unfolded dimeric intermediates (I₂) is reported [Blackburn and Noltmann, 1981]. The compact monomeric intermediate structure may not be identical to the conformation of the subunits in the native dimer, but some native-like secondary or tertiary folding may exist e.g. denaturation of superoxide dismutase [Mei et al., 1992], glutathione-S-transferase [Sacchetta et al., 1993], etc. Dimeric proteins can follow a 2-state transition, \( N₂ \leftrightarrow 2D \), where \( N₂ \) is native dimer and \( D \) is the denatured monomer. The 2-state model implies that the native monomer (or a monomeric intermediate) does not exist at significant concentration at equilibrium i.e. the quaternary interactions are necessary for stabilization of the folded monomeric state. Two-state denaturation of dimeric proteins have been reported for Arc repressor [Milla and Sauer, 1994], SIV.
protease [Grant et al., 1992], HIV protease [Grant et al., 1992], repressor of primer [Steif et al., 1993], neurotrophin-3 and neurotrophin-4/5 [Timm et al., 1994], etc. The midpoint of thermal [Pakula and Sauer, 1989] or chemical [Bowie and Sauer, 1989b] denaturation transition can also be computed from the denaturation curves and can be used to compare 2 different proteins or a single protein behaving differently under the effects of different denaturing conditions. There are various reports available which have shown different effects of GdnHCl and urea on protein unfolding like mushroom tyrosinase which shows different behaviour towards GdnHCl, urea and SDS in terms of different transition processes for these denaturants [Park et al., 2003]. Behaviour of human placental and goat lung cystatins towards these denaturants was also found to be different in terms of midpoint of transition and presence of intermediate states [Rashid et al., 2005; Khan and Bano, 2009b]. Increasing inclination in protein folding/unfolding is due to the recognition that failure of cellular protein folding mechanisms is associated with a variety of important human disorders ranging from cystic fibrosis to Alzheimer’s disease. There is a growing body of evidence indicating a critical role for partially folded protein conformers in the process of conversion of normal cellular proteins into disease causing, proteinase-resistant protein aggregates of various morphologies [Kelly, 1998; Dobson, 2004].

1.11 AMYLOID FIBRIL FORMATION

Amyloidogenesis is the aggregation of soluble proteins into structurally conserved fibers. Amyloid fibers are distinguished by their resistance to proteinase K and detergent, tinctorial properties and β-sheet rich secondary structure [Hammer et al., 2008]. Amyloid formation is a hallmark of many human diseases like Alzheimer’s [Gotz et al., 2009], diabetes mellitus [Engel et al., 2008], autosomal hereditary systemic amyloidosis [De Felice, 2004], prions diseases and more than 20 different human disorders like, HCCAA, Parkinson’s disease, Huntington’s disease, etc. [Merlini and Bellotti, 2003]. Amyloid fibres are incredibly stable β-sheet rich structures that many proteins can form [Smith et al., 2006; Holm et al., 2007]. The ability to fibrillate is independent of the original native structure of protein [Khurana et al., 2003] and overall yield and stability of the fibrils [Hortscansky et al., 2005]. This led Dobson and coauthors to propose that amyloid-fibril formation is a generic property of proteins [Dobson, 1999; Fandrich et al., 2001]. Fibrillation
generally starts from an intermediate state, either partially unfolded or partially folded, molten globule or native like intermediate [Rochet and Lansbury, 2000]. In case of globular proteins such as cystatin C [Ekiel and Abrahamson, 1996], stefin B [Zerovnik et al., 2007], partial unfolding and in case of unfolded polypeptides such as α-synuclein [Uversky et al., 2001] and islet amyloid peptide partial folding is must. In vitro, variation of solvent conditions by changing pH or adding organic solvents can lead to partial unfolding and subsequent protein fibril formation. With unfolded polypeptides, partial folding can be obtained by lowering pH or by heating. In vivo, partial unfolding may happen as a consequence of lowered protein stability due to mutation, local change in pH in membranes, oxidative and heat stress, whereas partial folding may happen on exposure to environmental hydrophobic substances, such as pesticides [Uversky et al., 2001].

**Amyloidogenic conformation and common structural traits of fibrils**

Ordered fibrillar aggregates and the amyloid-fibrils can be studied at lower resolution by transmission electron microscopy (TEM), atomic force microscopy (AFM) [Goldsbury et al., 1997; Ding and Harper, 1999], cryo-electron microscopy, X-ray diffraction and solid state NMR. Common features of the fibrils are [Serpell, 2000a], β-strands (separated by 4.7 Å) running perpendicular to the long axis of the fibrils and β-sheets extending parallel to the axis. The β-strands form a β-helical twist with usual repeat at every 115 or 250 Å [Serpell, 2000a]. There are two main types of fibrils, type 2 are built from two intertwined filaments, with a diameter from 80-130 Å. Type 1 fibrils are thinner and are formed from one filament only. There are other types of fibrils for e.g., a fibril and untwisted filaments of human stefin B [Zerovnik et al., 2002a]. The fibrils generally consist of 2-6 protofilaments, each ~2-5 nm in diameter, that generally twist together to form fibrils that are typically 7-13 nm wide [Serpell et al., 2000b]. The fibrils have the ability to bind specific dyes such as thioflavin T (ThT) and congo red (CR) [Krebs et al., 2005; Klunk et al., 1999].

**Kinetic basis of fibrillogenesis**

Fibrillogenesis often starts with dimers as building blocks which oligomerize to tetramers, octamers etc., constituting the prefibrillar aggregates composed of fluid nuclei [Lomakin et al., 1996]. From these, protofibrils grow upto 200 nm in length and are slightly curved [Lomakin et al., 1996]. All these species accumulate in the lag
phase which is followed by an exponential growth phase in which protofibrils merge into filaments. Fully made fibrils are then made from filaments added laterally or by end to end [Aggeli et al., 2001]. Thioflavin T fluorescence is generally used to follow kinetics of fibril formation [Sabate and Saupe, 2007]. Addition of preformed fibrils (seeding) is known to speed up the process of fibrillation [Jenko et al., 2004]. These events are depicted in figure 8. Several amyloidogenic proteins form domain swapped dimers like cystatin C [Janowski et al., 2001], human stefin A [Staniforth et al., 2001], stefin B [Skerget et al., 2009]. It has been proposed that domain swapped dimers could lead to higher oligomerisation and amyloid fibrillization [Liu et al., 2001].

**Fibrillogenic cystatins**

Cystatins are prone to form amyloids [Morgan et al., 2008; Turk et al., 2008]. Human cystatin C is highly amyloidogenic protein. The fibril formation is also known to occur in chicken cystatin [Staniforth et al., 2001], stefin B under in vitro conditions [Zerovnik et al., 2007], stefin A [Jenko et al., 2004], latexin [Pallares et al., 2007], CRES protein [von Horsten et al., 2007]. In the case of HCC the oligomers and fibrils are formed by propagated domain swapping. This model is not compatible with stefin B, in which proline (Pro) isomerization is important in preventing steric clashing [Morgan et al., 2008]. Trans to cis isomerization of Pro 74 is involved in formation of stefin B dimers. Since this Pro is widely conserved in cystatin superfamily its isomerization can play role in amyloidogenesis [Jenko-Kokalj et al., 2007].

**Protein fibrillation: connection to pathophysiology and disease**

About 20 human proteins have been found in proteinaceous deposits in various conformational diseases [Zerovnik, 2002]. There is absence of any sequence or structural homology in these proteins, but a common event is thought to be a conformational change leading to lack of biological function or gain of toxic activity, and possibly formation of amyloid fibrils. Co-localization of protein aggregates with degenerating tissue and association of their presence with disease symptoms indicate the involvement of amyloid deposition in the pathogenesis of conformational diseases [Soto, 2001]. Amyloid cytotoxicity appears to be associated with prefibrillar aggregated states either because of their ability to permealize cell membranes to general ion flux [Ceru et al.; 2008; Rabzelj et al., 2008] or because the rather diffuse hydrophobic surface may catalyze unwanted reactions [Bucciantini et al., 2004].
Fig. 8  Properties of amyloid polymerization

(A) A graphic representation of amyloid fiber polymerization displaying nucleus dependent kinetics (black line). Preformed amyloid fibers can act as seeds to speed the kinetics of fiber polymerization (grey line). This process eliminates the lag phase associated with nucleus formation.

(B) Model of amyloid fiber polymerization. A build up of monomer occurs which leads to the formation of multimers and finally the amyloid fiber end product. Large arrows represent processes that are energetically favourable while small arrows represent energetically unfavourable processes.

(A) Fiber Assembly

100%
75%
50%
25%

seeded
unseeded

Time

(B) Log Growth

Lag Phase

Monomer Dimer Oligomer nucleus Fiber
Anti-amyloidogenic and fibril destabilizing agents

A number of agents are employed with the aim to either inhibit or reverse the conformational change, or to dissolve the smaller aggregates and disassemble the amyloid fibrils. Few such approaches include ‘β-sheet breakers’ or ‘mini-chaperones’ [Soto, 2001], nicotine and melatonin [Findeis, 2000], apomorphine [Lashuel et al., 2002], various antibiotics [Zerovnik, 2002]. There are controversial reports on the inhibitory effects of metal ions on amyloid fibril formation. Recently, it was shown by Raman et al. [2005] and Zerovnik et al. [2006] that binding of Cu$^{2+}$ and Zn$^{2+}$ but not Fe$^{3+}$ to amyloid-β-peptide retards amyloid fibril formation and Cu$^{2+}$ binding to stefin B inhibits amyloid fibrillation. Contrarily, promotion of aggregation and fibrillation by presence of Cu$^{2+}$ has been shown for prion protein [Brown et al., 1997], α-synuclein [Rasia et al., 2005], amyloid-β-protein [Atwood et al., 1998]. Recently, several natural compounds, like polyphenols, curcumin etc., have been demonstrated to remarkably inhibit the formation of fibrillar assemblies in vitro and their associated cytotoxicity [Riviere et al., 2008].

1.12 REACTIVE SPECIES MEDIATED PROTEIN DAMAGE

Radical mediated damage to proteins, initiated by electron leakage, metal-ion-dependent reactions and auto-oxidation of lipids and sugars results in production of protein hydroperoxides and aggregation or fragmentation of proteins. Damaged proteins are often functionally inactive and their unfolding is associated with enhanced susceptibility to proteinases.

Reactive species

Free radicals such as reactive oxygen (ROS) and reactive nitrogen species (RNS) are well recognised for playing a dual role as both deleterious and beneficial species [Valko et al., 2007]. Important physiological functions that involve free radicals or their derivatives are summarized in Table 4. Besides these, free radicals are also involved in ATP generation, apoptosis of effete or defective cells, production of prostaglandins and leukotrienes etc. [Devasagayam et al., 2004].

Nature has endowed cells with protective antioxidant mechanisms to neutralize and eliminate the harmful reactive species. However, oxidative stress can occur when there is a disturbance in the pro-oxidant/antioxidant systems in favour of the former
Overproduction of RNS is termed as nitrosative stress [Ridnour et al., 2004]. The uncontrolled oxidative stress initiates a series of harmful biochemical events associated with diverse pathological processes [Juranek and Bezek, 2005]. Reactive species are small molecules with an oxygen or nitrogen atom in their structure [Halliwell and Gutteridge, 2007]. These ROS and RNS can be free radicals with an unpaired electron [e.g. 'NO (nitric oxide radical), O$_2^\cdot$ (superoxide radical) and OH$^\cdot$ (hydroxyl radical)] or non-radicals (e.g. H$_2$O$_2$). They can be anions [e.g. O$_2^-$ (superoxide) and ONOO$^-$ (peroxynitrite)] or non-ions (e.g. H$_2$O$_2$, 'NO, OH').

**Superoxide (O$_2^\cdot$~)**

It is relatively innocuous, however, its reaction with other radicals like 'NO and iron clusters in some of the enzymes makes it the mother of potent reactive species collectively called as ROS. The mitochondrial respiratory chain is the most important site of O$_2^\cdot$~ generation [Turrens, 2003]. In biological tissues O$_2^\cdot$~ can be converted non-enzymically into non-radical species H$_2$O$_2$ and singlet oxygen [Steinbeck et al., 1993]. The electronic ground state of dioxygen is a triplet ($^3$O$_2$), having two unpaired electrons with same spin and hence a diradical, and a one-electron poor oxidant. Pairing these electrons in opposite spins give singlet oxygen ($^1$O$_2$) a two electron potent oxidant. Singlet oxygen is produced in several physiological processes for e.g. by reaction of H$_2$O$_2$ and hypochlorous acid (HOCI) in neutrophils,

\[
H_2O_2 + HOCl \rightarrow ^1O_2 + H_2O + HCl
\]

ROS has been implicated in etiology of various diseases [Droge, 2002; Halliwell and Gutteridge, 2007].

**Riboflavin**

It is an important constituent of our daily diet, present in free and conjugated forms in almost all biological tissues and fluids [Spector, 1980; Rose et al., 1986]. Flavins are known to photooxidize amino acids and effect the conformation of proteins [Joshi, 1985; Baba et al., 2004]. They are known to generate $^1$O$_2$, OH$^\cdot$, flavin triplet state, O$_2^\cdot$~ and H$_2$O$_2$ which can modify proteins and other biological macromolecules [Keynes et al., 2003; Cardoso et al., 2006].
Hydrogen peroxide

It is produced continuously in all cells and is often employed as a signalling molecule. Although not a free radical, it has a great physiological relevance because of its ability to penetrate biological membranes and to act like an intermediate in the production of more reactive oxygen species, namely hydroxyl radical and hypochlorous acid [Nordberg and Arner, 2001]. It does not readily oxidize most proteins, lipids or DNA but is cytotoxic at micromolar concentrations and has been implicated in number of diseased states [Pryor et al., 2006].

Nitric oxide (\(^{•}\text{NO}\))

It is a free radical endogenously produced in a variety of mammalian cells by both constitutive and inducible forms of nitric oxide synthase [Ghafourifar and Cadenas, 2005]. \(^{•}\text{NO}\) is an important mediator of a variety of diverse biochemical and physiological processes, like signal transduction, neurotransmission, smooth muscle relaxation, platelet inhibition, blood pressure modulation, immune system control, macrophage mediated cytotoxicity [Blaise et al., 2005]. Reactivity of \(^{•}\text{NO}\) as a free radical species is quite weak which combined with its lipophilicity allows it to be remarkably diffusible [Denicola et al., 2002]. In vivo, this diffusion is largely regulated by its reaction with haemoglobin [Liu et al., 1998]. \(^{•}\text{NO}\) is a mother to a family of reactive compounds, collectively called RNS [Table 5]. Their production and pathophysiological effects are represented in Fig. 9. \(^{•}\text{NO}\) can also react with \(\text{O}_2^{-}\) to give peroxynitrite. \(^{•}\text{NO}_2\) can also react with \(\text{O}_2^{-}\) giving peroxynitrate, a more powerful oxidant than peroxynitrite. Biological oxidations by peroxynitrate could result either directly or by its decomposition products [Pryor et al., 2006].

\[
\begin{align*}
^{•}\text{NO}_2^{-} & \quad + \quad \text{O}_2^{-} & \quad \rightarrow & \quad \text{O}_2\text{NOO}^{-}^{-} \\
(\text{Nitrogen dioxide}) & \quad (\text{Superoxide radical}) & \quad (\text{Peroxynitrate}) \\
\text{O}_2\text{NOO}^{-}^{-} & \quad \rightarrow & \quad ^1\text{O}_2 & \quad + \quad \text{NO}_2^{-} \\
(\text{Peroxynitrate}) & \quad & \quad (\text{Singlet oxygen}) & \quad (\text{Nitrate})
\end{align*}
\]

Peroxynitrite can react with carbon dioxide generating nitrosoperoxycarboxylate (NPC), another potent protein damaging agent [Tien et al., 1999]. Overproduction of \(^{•}\text{NO}\) can mediate toxic effects, e.g., DNA fragmentation, cell damage and neuronal
TABLE 4: IMPORTANT PHYSIOLOGICAL FUNCTIONS THAT INVOLVE FREE RADICALS OR THEIR DERIVATIVES

<table>
<thead>
<tr>
<th>Type of Radical</th>
<th>Source of Radical</th>
<th>Physiological Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nitric oxide ('NO)</td>
<td>Nitric oxide synthase</td>
<td>Smooth muscle relaxation (control of vascular tone) and various other cGMP dependent functions</td>
</tr>
<tr>
<td>Superoxide ((\text{O}_2^-)) and related ROS</td>
<td>NAD(P)H oxidase</td>
<td>Control of ventilation Control of erythropoietin production and other hypoxide inducible functions Smooth muscle relaxation Signal transduction from various membrane receptors/enhancement of immunological functions</td>
</tr>
<tr>
<td>Superoxide ((\text{O}_2^-)) and related ROS</td>
<td>Any Source</td>
<td>Oxidative stress responses and the maintenance of redox homeostasis</td>
</tr>
</tbody>
</table>

*Adopted from Droge [2002].

TABLE 5: NITROGEN OXIDES IMPLICATED IN THE BIOLOGICAL SEQUEL OF NITRIC OXIDE FORMATION

<table>
<thead>
<tr>
<th>Nitrogen Oxide</th>
<th>Common Name</th>
<th>Reactivity and Reaction Types</th>
<th>Substrates</th>
</tr>
</thead>
<tbody>
<tr>
<td>(^{1}\text{NO})</td>
<td>Nitric Oxide</td>
<td>Weak nitrosylation</td>
<td>Transition metals, radicals, oxygen</td>
</tr>
<tr>
<td>(\text{NO}^+)</td>
<td>Nitrosonium</td>
<td>Moderate nitrosation</td>
<td>Thiols, amines</td>
</tr>
<tr>
<td>(\text{NO}^-)</td>
<td>Nitroxyl</td>
<td>Strong oxidation</td>
<td>Thiols, lipids, metals, oxygen, DNA</td>
</tr>
<tr>
<td>(\text{NO}_2)</td>
<td>Nitrogen Dioxide</td>
<td>Strong oxidation, nitrination</td>
<td>Antioxidants, thiols, lipids</td>
</tr>
<tr>
<td>(\text{N}_2\text{O}_3)</td>
<td>Dinitrogen Trioxide</td>
<td>Strong nitrosation, oxidation</td>
<td>Thiols, amines, lipids, antioxidants</td>
</tr>
<tr>
<td>(\text{ONOO}^-)</td>
<td>Peroxynitrite</td>
<td>Strong nitration, nitrosation, oxidation</td>
<td>Lipids, tyrosine, phenylalanine, DNA, thiols, antioxidants</td>
</tr>
<tr>
<td>(\text{NO}_2^+)</td>
<td>Nitronium</td>
<td>Strong nitration</td>
<td>Tyrosine, amines, phenylalanine</td>
</tr>
</tbody>
</table>
Fig. 9  A schematic diagram of the conversion of nitric oxide (NO) to other nitrogen oxides

Hydrophilic interior of membranes will act as a lens to magnify the oxidation of NO to NO$_2$ or N$_2$O$_3$. The profound reactivity of these nitrogen oxides, as well as the formation of the highly cytotoxic peroxynitrite (ONOO$^-$), likely mediates cellular injury and tissue dysfunction in response to augmented NO production.

PARP (poly-ADP ribose polymerase); MMP (matrix metalloproteinase).
2NO + O₂ → 2NO₂

Lipid Membrane

NO

O₂

2NO + O₂ → 2NO₂

N₂O₃

N₂O₃

NO₂⁻

NO₂⁻

NO + NO₂⁻ → ONOO⁻

ONOO⁻

CO₂

CO₃⁻

OH⁻

NO₂⁻

NITROSATION
- SH
- NH₂

• Protein modification

• PARP activation
• MMP activation
• DNA strand break
• Protein nitration
• - SH oxidation

Tissue injury

• apoptosis
• necrosis
cell death [Dawson et al., 1992]. NO also shows neurotoxicity and act as pathological mediator in cerebral ischemia, epilepsy, Alzheimer’s disease, Parkinson’s disease and certain neurodegenerative diseases [Moncada et al., 1991] and it is also involved in toxicity associated with diabetes [Pacher et al., 2005] and pancreatitis [Chvanov et al., 2005].

**Hypochlorous acid (HOCl)**

One of the strongest physiological oxidants known, is produced in vivo at inflammation sites by enzymatic oxidation of chloride ions,

\[
\text{H}_2\text{O}_2 + \text{Cl}^- \xrightarrow{(\text{MPO})} \text{HOCl} + \text{OH}^-
\]

(Hydrogen (Chloride ion) (Hypochlorous (Hydroxyl ion)
peroxide) acid)

Where, MPO is enzyme myeloperoxidase.

It is estimated that between 25 and 40% of H_2O_2 generated by activated neutrophils is used to form HOCl [Babior, 2000]. An additional reaction of HOCl is with nitrite (NO_2^-) to form nitryl chloride (NO_2Cl), a reaction favoured with decreasing pH [Eiserich et al., 1998],

\[
\text{HOCl} + \text{NO}_2^- + \text{H}^+ \rightarrow \text{NO}_2\text{Cl} + \text{H}_2\text{O}
\]

(Hypochlorous (Nitrite) (Nitryl chloride) (Water)
acid)

HOCl is a reactive chlorine species (RCS) capable of chlorinating protein tyrosine and oxidising many important biomolecules like DNA, collagen and ATPase, etc. and causing cell death [Winterbourn and Kettle, 2000; Jenner et al., 2002]. Nitryl chloride is capable of nitrating, chlorinating and dimerizing phenol compounds such as tyrosine [Eiserich et al., 1998].

**Protein modification by reactive species**

Free amino acids and amino acid residues in proteins are highly susceptible to oxidation by one or more reactive species (ROS/RNS/RCS) that (a) are present as pollutant in atmosphere (b) are generated as by products of normal metabolic processes and (c) are formed during exposure to X, γ, or UV radiations. Studies have shown that oxidation of proteins can lead to hydroxylation of aromatic groups and aliphatic amino acid side chains, nitration of amino acid residues, nitrosylation of sulphhydril groups, sulphoxidation of methionine residues, chlorination of aromatic
and primary amino groups, and conversion of some amino acid residues to carbonyl
derivatives [Stadtman and Levine, 2003; Dean et al., 1997; Yan and Sohal, 2002;
Davies et al., 1987a; 1987b; Peskin and Winterbourn, 2001]. The oxidation of
proteins by reactive species can also lead to the cleavage of peptide bonds [Garrison,
1987; Uchida et al., 1990].

Exposure of proteins to free radicals causes (a) protein inactivation, e.g., of catalase
and Mn superoxide dismutase by \( \cdot \text{NO} \) [Sigfrid et al., 2003; Castro et al., 2004], \( \alpha_2 \)-
macroglobulin by HOCl and H\(_2\)O\(_2\) [Khan and Khan, 2004], goat lung cystatin by \( \cdot \text{NO} \)
[Khan et al., 2009] and sheep plasma kininogen by ROS [Baba et al., 2004], etc. (b)
protein aggregation, fragmentation or cross linking [Davies and Delsignmore, 1987;
Hawkins and Davies, 1998; Chapman et al., 2003; Verzyl et al., 2000; Di Mascio et
al., 2000].

Modestly oxidized proteins are usually more sensitive to proteolytic attacks by most
proteinases [Davies et al., 1987b; Wolff and Dean, 1986] whereas heavily oxidized
proteins have decreased susceptibility [Davies et al., 1987b] and can be a
cause/consequence of certain diseased states like aging, atherosclerosis and
neurodegeneration.

**Vulnerability of pancreas to free radical damage**

Pancreas is highly susceptible to free radical attack. Enzymatic antioxidant defence
mechanisms of pancreatic \( \beta \)-cells are particularly weak [Cobianchi et al., 2008;
Lenzen, 2008a] and can be overwhelmed by redox imbalance arising from
overproduction of ROS/RNS. The consequence of this redox imbalance are lipid
peroxidation, protein oxidation, DNA damage and interference of reactive species
with signal transduction pathways which contribute significantly to \( \beta \)-cell dysfunction
and death in Type 1 and Type 2 diabetes mellitus [Pacher et al., 2005; Lenzen, 2008b].
Quite similar is the case with acinar cells, mainly involved in the secretion and
synthesis of digestive enzymes. There are numerous indications that ROS/RNS play a
significant role in chronic and acute pancreatitis [Rau et al., 2000; Sandstrom et al.,
2005; Shimizu, 2008]. Primary injury to acinar cells results in intracellular
trypsinogen activation and inhibition of acinar cell secretion followed by ROS
production leading to damage of biomolecules, membranes and activation of
inflammatory cells which produce more ROS/RNS, responsible for acinar necrosis
and amplification of the inflammation in pancreas. Furthermore, zymogen granules,
the most abundant organelle in acinar cells lack full set of scavenger characteristics of intact cells [Niederau et al., 1996] hence are liable to oxidative damage. The structural and functional impairment of these granules leads to leakage of trypsin and cell damage enhancement. Oxidative and nitrosative stress [Ischiropoulos et al., 2003; Ghafourifar et al., 1999] has been documented in pancreatic tissue by detection of ROS generation [Urunuela et al., 2002] and accumulation of products of ROS-mediated lipid peroxidation, protein oxidation and depletion of low molecular weight antioxidants [Rau et al., 2000].

**Antioxidative and antinitrosative stress activities of caffeic acid, quercetin and curcumin: protection against free radicals**

Knowledge about prevention of protein oxidation is cursory. Certain approaches attenuate the secondary radicals and confer cytoprotection [e.g. mesalamine against ONOO⁻] but fail to affect the radical sinks [Sandoval et al., 1997]. Antioxidant agents of natural origin like polyphenols, flavonoids etc. have attracted special interest lately, because of their high efficacy and multifaceted health benefits [Ullah and Khan, 2008]. Several such natural components have proven to be beneficial in various disorders of pancreas like, curcumin against islet cell damage [Meghana et al., 2007; Kanitkar et al., 2008], quercetin against beta-cell damage [Kim et al., 2007; Coskun et al., 2005] and caffeic acid against pancreatic damage by free radicals [Lapidot et al., 2002].

**Caffeic acid (3, 4-dihydroxycinnamic acid)**

Caffeic acid (CA) [Fig. 10 (C)] is a naturally occurring phenolic compound found in many fruits, vegetables, and herbs including coffee [Gulcin, 2006]. It has antioxidant, anti-ischemia reperfusion, antithrombosis, antihypertension, antifibrosis, antivirus, antitumor and antidiabetic activities. It is known to quench superoxide and hydroxyl radicals, and RNS [Gulcin, 2006; Olmos et al., 2008; Takahama et al., 2009].

**Quercetin (3, 3’, 4’, 5, 7-pentahydroxyflavone)**

Quercetin [Fig. 10 (B)] is found bound to one or two glucose molecules (monoglycoside and diglycoside forms) as one of the most abundant dietary flavonoids in apples; black, green and buckwheat tea; onions; raspberries; redgrapes;
citrus fruits and other green leafy vegetables [Hertog and Hollman, 1996]. It has been shown in vitro to act as an antioxidant [Filipe et al., 2004], inhibit nitric oxide pathway [Mu et al., 2001], and have anti-inflammatory and anticancer activities [Wadsworth et al., 1999; Mertens-Talcott and Percival, 2005].

Quercetin and caffeic acid bear structural groups, responsible for direct scavenging of free radicals: o-dihydroxy structure in caffeic acid and in ‘B’ ring of quercetin and the 2, 3-double bond in conjugation with the 4-oxo function in the ‘C’ ring and 3- and 5-hydroxyl groups with 4-oxo function in the ‘A’ and ‘C’ rings of quercetin [Fig. 10].

Curcumin (diferuloyl methane)

Curcumin a bioflavonoid is the colouring pigment present in rhizomes of Curcuma longa. Curcumin has been shown to exhibit antioxidant, anti-inflammatory, antiviral, antibacterial, antifungal and anticancer activities and thus has a potential against various malignant diseases, diabetes mellitus, allergies, arthritis, Alzheimer’s and other chronic illnesses [Sreejayan and Rao, 1997; Aggarwal et al., 2007; Dhillon et al., 2008]. Its polyphenolic structure with two ferulic acids linked via methylene bridge at the C atoms of the carboxyl groups [Fig. 10 (A)], double bonds in the alkene part of the molecule, hydroxyl groups of the benzene ring and central β-diketone allows it to directly scavenge free radicals like nitric oxide, ROS etc. [Sreejayan and Rao, 1997].

1.13 DRUG-PROTEIN INTERACTION: EFFECT OF PANCREATITIS CAUSING AND ANTIDIABETIC AGENTS

The binding and interaction of drugs with plasma and tissue proteins strongly affects their distribution, metabolism as well as pharmacodynamics and toxic properties. Accumulation of drug molecules at certain sites in the body causing a localized high concentration, adverse drug reactions [Wen and Ye, 1993] and ligand induced protein structure conformational changes [Takeda et al., 1988] are major problems complicating drug medical therapy. Therefore, studies analyzing the binding mechanism between proteins and drugs and the structure of resulting complexes are of particular interest. These works enable to elucidate how ligand affinity is regulated and how the protein conformation is altered upon complexation. Acute pancreatitis, an autodigestive disorder, typically presents as an acute inflammation of the pancreas.
Fig. 10  Chemical structures of A) Curcumin B) Quercetin C) Caffeic acid
(A) Curcumin

(B) Quercetin

(C) Caffeic Acid
Gallstones and heavy alcohol are major causes of this condition besides hypertriglyceridemia, hyperparathyroidism, pancreatic tumors and surgery [Greenberger and Toskes, 2006]. Many frequently prescribed drugs are suspected to cause acute pancreatitis (AP), referred to as drug induced pancreatitis (DIP), accounting for at least 2-5% of reported cases of AP. Drugs commonly associated with pancreatitis belong to several classes, like antimicrobials, anti-inflammatory, antineoplastic, immunomodulating and cardiovascular agents and antiepileptic drugs (valproic acid, marketed generally as its sodium salt, sodium valproate) [Trivedi and Pitchumoni, 2005]. The mechanism of DIP is not clear. However, it may be caused by direct toxicity of drugs or by drug induced indirect mechanisms like ischemia, intravascular thrombosis and increased viscosity of pancreatic juices. One hallmark of AP is cathepsin B mediated trypsinogen activation [Halangk and Lerch, 2000] and oxidative/nitrosative stress [Shimizu, 2008]. Valproic acid has been categorised as class I medications associated with AP [Trivedi and Pitchumoni, 2005].

**Valproic acid**

Valproic acid (VPA, 2-propylpentanoic acid/ Sodium 2-propyl pentanoate), an 8-carbon branched chain fatty acid [Fig. 11 (A)] is an established drug for the treatment of epileptic seizures and mania in bipolar disorder [Bowden and Singh, 2005]. VPA is an emerging anticancer drug too [Kostrouchova et al., 2007]. However, it is one of most incriminate drugs causing AP besides being classified as teratogen [Norgaard et al., 2006; Gerstner et al., 2007; Werlin and Fish, 2006]. Other side effects of VPA include fatal hepatotoxicity, hyperammonenmic encephalopathy and coagulation disorders [Gerstner et al., 2007]. No relationship between the occurrence of pancreatitis and duration of VPA therapy, dosage and serum level has been documented [Werlin and Fish, 2006]. Research indicates that VPA's cytotoxic activity is the result of generation of hydrogen peroxide and production of highly reactive hydroxyl radicals [Graf et al., 1998; Kawai and Arinze, 2006].

Antidiabetic drugs may be subdivided into six groups, insulin, sulfonylureas, alphaglucosidase inhibitors, biguanides, meglitinides and thiazolidinediones. Diabetes is characterized by insufficiency in insulin secretion and/or action. Increased production of ROS/RNS is also observed in diabetes [Lenzen, 2008a; 2008b]. Since
Fig. 11 Chemical structures of A) Valporic acid B) Metformin hydrochloride C) Glimepiride
(A) Valproic Acid

(B) Metformin

(C) Glimepiride
the introduction of insulin in the 1920s as a treatment for diabetes [Banting and Best, 1990], it is considered to be the most effective treatment.

**Metformin hydrochloride**

Metformin hydrochloride [N, N'-dimethylimidodicarbonimidic diamide hydrochloride, Fig. 11 B] is the only member of biguanide class of antihyperglycemics prescribed with increasing frequency for treatment of Type 2 diabetes. Its overdosage and chronic use is related to incidences of lactic acidosis.

**Glimepiride**

Glimepiride [Fig. 11 C] is a hypoglycemic agent of sulfonylurea class. It acts by increasing insulin secretion from pancreas and insulin sensitivity in peripheral cells. Sulfonylureas (glimepiride) cause β-cell apoptosis, produce ROS and their prolonged exposure causes disturbances in islet cell function [Sawada et al., 2008; Del Guerra et al., 2005].

1.14 SCOPE OF THE THESIS

Intracellular protein degradation occurs in two major cellular systems: lysosomal and non-lysosomal ubiquitin-proteasome systems. In the lysosomal pathway, protein degradation is a result of combined random and limited action of cathepsins. Recently, a host of functions are annexed to cathepsins, like antigen processing and presentation, bone remodelling, apoptosis mediation, etc. [Turk and Turk, 2008]. Potentially highly destructive activity of cathepsins can be regulated by their endogenous protein inhibitors, members of cystatin superfamily. A balance between proteolytic activity and proteinase inhibition is imperative to the appropriate functioning of many biological processes. Also, TPIs have been accredited with a multitude of roles and their non- or dys- functional states have been incriminated in various pathological conditions. This has stipulated the purification and characterization of members of cystatin superfamily from varied sources, since the first report of such an inhibitor from chicken egg white [Fossum and Whitaker, 1968]. Pancreas is one of the most complex tissue of the mammalian body and also capital of proteolytic power. It functions to maintain the glucose homeostasis of the body and participates in digestion. However, regulation of pancreatic cathepsins by proteinaceous inhibitors
remains a less ventured area.

Keeping this in view, the specific objectives of the planned research included:

**Chapter 1**

Realizing the role of cathepsins in pancreas the importance of thiol proteinase inhibitors (TPI) in their regulation was envisaged. Isolation and partial purification of TPI from pancreas of goat was performed by gel filtration chromatography. Its detailed biochemical characterization was undertaken, which included elucidation of its molecular weight, hydrodynamic properties, pH and thermal stability, partial amino acid sequence. Its interaction with model cysteine proteinase, papain was worked out to determine its kinetic properties.

**Chapter 2**

To assess the conformational stability of the isolated inhibitor, equilibrium denaturation studies were conducted with classical denaturants, guanidine hydrochloride and urea employing fluorescence and CD spectroscopy.

**Chapter 3**

Cystatins are predisposed to form amyloids. Certain pancreatic proteins are also known to undergo fibrillation (e.g. amylin). Thus, experiments were conducted to judge the propensity of isolated TPI for fibrillation under various conditions (pH variance, presence of organic solvent, trifluoroethanol, TFE) employing Transmission electron microscopy and Thioflavin T fluorescence assay. The possibility of disintegration of preformed fibrils and inhibition of de novo fibrillation of the isolated inhibitor by divalent metal cations, Zn$^{2+}$ and Cu$^{2+}$, was also studied.

**Chapter 4**

Proteins are most liable to free radical damage because of their abundance. Pancreas also has high proclivity of oxidative and nitrosative stress build up because of its feeble antioxidant defence. Furthermore, ROS and RNS are recognized as major culprits in pancreatitis, β-cell damage in diabetes and pancreatic stellate cell fibrosis. Thus affect of various reactive species (a) ROS (b) •NO (c) H$_2$O$_2$ (d) HOCl, was determined on activity and structure of the isolated inhibitor by fluorescence
spectroscopy and PAGE. Natural polyphenols and flavonoids, caffeic acid, curcumin and quercetin, offering inexpensive and non-toxic modes of anti-free radical therapy were analyzed to ascertain their efficacy and efficiency in preventing ROS/nitric oxide and hypochlorous acid induced damaged to the isolated inhibitor.

Chapter 5

Drug-protein interactions are determining factors in the therapeutic, pharmacodynamic and toxicological drug properties. Experiments were conducted to find out the effects of pancreatitis causing sodium valproate and antidiabetic agents (insulin, metformin, glimepiride) on structure and function of the purified inhibitor employing UV, Fluorescence and CD spectroscopy and PAGE. The results of the present studies showed the presence of a 44 kDa thiol proteinase inhibitor in goat pancreas subsequently named as PTPI. It was found to be highly specific and efficient inhibitor of CPs. Its physical, kinetic and biochemical properties commensurate with the cystatin superfamily signatures. The likelihood of its being amyloidogenic was proved by transmission electron microscopy (TEM) and ThT fluorescence. PTPI was found vulnerable to ROS, nitric oxide and hypochlorous acid. However, curcumin, quercetin and caffeic acid protected the free radical inflicted structural and functional alterations. Complexation with drugs caused significant conformational and functional changes in PTPI. Conclusively, these results forestall the significance of PTPI in vivo, in pancreatic (patho) physiology.