The genetic and molecular bases of monogenic disorders affecting proteolytic systems

I Richard

Complete and limited proteolysis represents key events that regulate many biological processes. At least 5% of the human genome codes for components of proteolytic processes if proteases, inhibitors, and cofactors are taken into account. Accordingly, disruption of proteolysis is involved in numerous pathological conditions. In particular, molecular genetic studies have identified a growing number of monogenic disorders caused by mutations in protease coding genes, highlighting the importance of this class of enzymes in development, organogenesis, immunity, and brain function. This review provides insights into the current knowledge about the molecular genetic causes of these disorders. It should be noted that most are due to loss of function mutations, indicating absolute requirement of proteolytic activities for normal cellular functions. Recent progress in understanding the function of the implicated proteins and the disease pathogenesis is detailed. In addition to providing important clues to the diagnosis, treatment, and pathophysiology of disease, functional characterisation of mutations in proteolytic systems emphasises the pleiotropic functions of proteases in the body homeostasis.

INTRODUCTION

Proteases catalyse the hydrolysis of peptide bonds between two adjacent amino acids. This irreversible reaction is widely used to control biological events both inside and outside the cell. It operates to eliminate damaged proteins, control protein levels, present antigenic peptides, activate precursors, or alter an existing function of a protein. Proteases are therefore key signaling proteins in many processes such as embryonic development, coagulation, immunity, cell differentiation, and cell death.

Proteases are divided into five different groups depending on the type of catalytic reaction they mediate (serine, cysteine, threonine, aspartate, and metallo proteases). They are further subdivided according to the structure of their catalytic sites into clans and families. To date, 504 sequences for proteolytic enzymes are reported for the human genome in the MEROPS database (http://merops.sanger.ac.uk). In addition, 182 non-proteolytic homologues have been identified. Many more genes are involved in proteolysis if the other participants in proteolytic events, such as inhibitors, cofactors or substrates, are taken into account. In particular, it is worth mentioning the very abundant class of E3 ubiquitin ligases, enzymes that determine the specificity of and timing for proteasomal degradation of target proteins.

Mutations in numerous proteases have been implicated in the aetiology of human disorders. Pathogenesis can be caused by uncontrolled activation of proteases resulting from specific gain of function mutations. When a mutation on one allele produces a level of activated proteases sufficient to generate a phenotype, it presents a dominant transmission. Conversely, pathogenesis can be caused by inactivation of the proteolytic activity. The phenotype will then be transmitted in a dominant fashion if the mutation produces a dominant negative form of the protease or in cases of haploinsufficiency. A recessive transmission mode is obtained in case of true loss of function mutations. In this case, half the normal enzyme is enough to fulfil the function, and a phenotype arises only when no activity is present in homozygotes.

The purpose of this review is to provide insights into the current knowledge about the molecular genetic causes of disorders of the proteolytic systems. Recent progress in understanding the function of the implicated proteins and their disease pathogenesis are also presented. This review focuses on monogenic
disorders with mutations in genes coding for proteases, describes the pleiotropy and emphasises the importance of the functions of proteases in the body homeostasis.

**GENETIC DISORDERS ASSOCIATED WITH SERINE PROTEASES**

Serine proteases are a class of abundant proteases in humans. They intervene in proteolytic cascades in the digestive tract and plasma and in various processes occurring in the secretory pathway and at transmembrane level (fig 1A, table 1). The human genetic disorders implicating this class of proteases parallel the diversity of their functions, and can result in problems with digestion, coagulation, complement activation, and impairment in synaptic plasticity, channel activity, prohormone maturation, stem cell differentiation, signalling, and degradation of substrates within the lysosome (fig 1B).

**Digestion**

The main part of protein digestion is achieved by pancreatic proteases, mainly trypsin (existing as cationic and anionic isoforms) and chymotrypsin. They are synthesised by exocrine cells as the inactive proenzymes trypsinogen and chymotrypsinogen, and are packaged into secretory vesicles together with trypsin inhibitor. Once released into the lumen of the small intestine, trypsinogen is activated to trypsin by enterokinase, a transmembrane enterocyte protease. Trypsin, in turn, activates chymotrypsinogen. Mutations in genes participating to this cascade have been identified in humans. Gain of function mutations in cationic trypsinogen lead to an autosomal dominant form of chronic pancreatitis, which

---

**Figure 1** Disorders involving serine proteases. (A) Subcellular localisation of the different proteases involved; (B) the organs presenting the prominent phenotype.
begins with recurrent episodes of acute pancreatitis and progresses to exocrine and endocrine pancreatic insufficiency. These manifestations correspond to an autodigestion of the pancreas due to premature intrapancreatic protease activation of the mutant cationic trypsinogen or to an increased transactivation of anionic trypsinogen.

Loss of function mutations in enterokinase cause congenital enteropeptidase deficiency with retardation of growth during childhood, chronic diarrhoea, and generalised oedema. This severe protein malabsorption is prominent during childhood and corresponds to a lack of enterokinase dependent activation of digestive enzymes.

**The coagulation cascade**

Coagulation is activated in response to a rupture of the vascular epithelium (extrinsic pathway) or by endogenous conditions in the vessels (intrinsic pathway), resulting in platelet aggregation, activation of the coagulation cascade, and finally the formation of fibrin and a clot. The coagulation cascade corresponds to a chain of proteolytic events performed by the procoagulant factors VII, IX, X, XI, and XII, and thrombin, in association with cofactors V and VIII. Several control factors, such as antithrombin, proteins C and S, and plasmin, function as anticoagulants, ensuring that the coagulation is limited to the site of interest. Disturbance in the balance of coagulation can lead either to bleeding disorders or to thrombophilies.

Bleeding disorders present themselves as bleeding episodes in mucosa, joints, muscles, brain, and internal organs. They are mainly due to loss of function mutations in the serine proteases of the coagulation cascade, and include mutations in factor VII, IX (haemophilia B), X, XI (haemophilia C), or XII or in prothrombin. With the exception of haemophilia B, which is X linked, they are all transmitted according to autosomal recessive inheritance but with possible manifestations in heterozygotes. In addition, while not affecting a protease per se, there are also loss of function mutations in protease cofactor, such as factor VIII (X linked haemophilia A) and factor V (autosomal recessive).

Thrombophilies are hereditary predispositions to venous thrombosis and can arise by virtue of specific gain of function mutations in procoagulant proteins or loss of function mutations in genes encoding control proteins of the coagulation cascade.

Disorders involving miscellaneous serine proteases

Loss of function mutations in neurotrypsin, a brain serine protease, lead to a rare non-syndromic learning disabilities with autosomic recessive transmission. Patients have cognitive impairment with IQ <50. Neurotrypsin is secreted in mucosa, joints, muscles, brain, and internal organs. They are mainly due to loss of function mutations in the serine proteases of the coagulation cascade, and include mutations in factor VII, IX (haemophilia B), X, XI (haemophilia C), or XII or in prothrombin. With the exception of haemophilia B, which is X linked, they are all transmitted according to autosomal recessive inheritance but with possible manifestations in heterozygotes. In addition, while not affecting a protease per se, there are also loss of function mutations in protease cofactor, such as factor VIII (X linked haemophilia A) and factor V (autosomal recessive).

**Complement activation**

Complement is a major part of the innate immune system, and also participates in the adaptive immune response. Complement can be activated according to three different pathways: classic, mannose binding lectin, or alternative. These pathways correspond to a series of proteolytic events, sequentially activating proteins of the complement and ultimately producing a key enzyme called C3 convertase. A common final pathway follows the generation of C3 convertase, leading to the formation of the membrane attack complex that destroys microorganisms. Several genetic conditions involving the serine proteases of the cascades have been identified, all of which correspond to loss of function mutations, but with various immune consequences.

In the classic pathway, deficiencies in the serine proteases C1r, C1s and C2 have been described to lead to multiple autoimmune features, especially systemic lupus erythematosus, an autoimmune manifestation affecting skin, joints, and other organs. To reconcile the deficiencies in activator proteases with what appears to be a hyperactivation of the complement, the “waste disposal” hypothesis has been proposed. In this hypothesis, the complement system fails to eliminate cells that have undergone apoptosis, and the partially degraded components of those cells induce an autoimmune response.

In the lectin pathway, one case of mannan binding lectin serine protease 2 (MASP2) deficiency has been reported. MASP2 is recruited to form the enzymatically active mannan binding lectin–MASP complex, and thus generating C3 convertase. The homozygous nonsense mutation of this patient prevents MASP2 binding to lectin and leads to complement deficiency. The disorder manifests itself as a susceptibility to infections and chronic inflammatory signs.

In the alternative pathway, inherited recessive deficiency of complement factor I (CFI), a C3 inactivating protease, have been described in two families. The deficiency causes an uncontrolled activation of the alternative pathway resulting in secondary C3 depletion. Consequently, patients with CFI deficiency present a propensity to acquire opportunistic infections.

Disorders involving miscellaneous serine proteases

Loss of function mutations in neurotrypsin, a brain serine protease, lead to a rare non-syndromic learning disabilities with autosomic recessive transmission. Patients have cognitive impairment with IQ <50. Neurotrypsin is secreted in mucosa, joints, muscles, brain, and internal organs. They are mainly due to loss of function mutations in the serine proteases of the coagulation cascade, and include mutations in factor VII, IX (haemophilia B), X, XI (haemophilia C), or XII or in prothrombin. With the exception of haemophilia B, which is X linked, they are all transmitted according to autosomal recessive inheritance but with possible manifestations in heterozygotes. In addition, while not affecting a protease per se, there are also loss of function mutations in protease cofactor, such as factor VIII (X linked haemophilia A) and factor V (autosomal recessive).

Thrombophilies are hereditary predispositions to venous thrombosis and can arise by virtue of specific gain of function mutations in procoagulant proteins or loss of function mutations in genes encoding control proteins of the coagulation cascade. The gain of function mutations, transmitted in an autosomal dominant manner, have a high prevalence (2–5% in white populations) possibly due to a selective advantage of women during delivery. Only one case corresponds to a serine protease: the G20210A mutation of the 3’ untranslated region of the prothrombin gene, which is associated with a 30% increase in its concentration and activity. The loss of function mutations are rare, with fewer than 1/300–1/2000 individuals affected, and include deficiencies of the serine proteases, protein C, tissue plasminogen activator, and plasminogen.

In the classic pathway, deficiencies in the serine proteases C1r, C1s and C2 have been described to lead to multiple autoimmune features, especially systemic lupus erythematosus, an autoimmune manifestation affecting skin, joints, and other organs. To reconcile the deficiencies in activator proteases with what appears to be a hyperactivation of the complement, the “waste disposal” hypothesis has been proposed. In this hypothesis, the complement system fails to eliminate cells that have undergone apoptosis, and the partially degraded components of those cells induce an autoimmune response.

In the lectin pathway, one case of mannan binding lectin serine protease 2 (MASP2) deficiency has been reported. MASP2 is recruited to form the enzymatically active mannan binding lectin–MASP complex, and thus generating C3 convertase. The homozygous nonsense mutation of this patient prevents MASP2 binding to lectin and leads to complement deficiency. The disorder manifests itself as a susceptibility to infections and chronic inflammatory signs.

In the alternative pathway, inherited recessive deficiency of complement factor I (CFI), a C3 inactivating protease, have been described in two families. The deficiency causes an uncontrolled activation of the alternative pathway resulting in secondary C3 depletion. Consequently, patients with CFI deficiency present a propensity to acquire opportunistic infections.
may lead to an excess of proteolytic cleavage of LDLR, preventing LDLR mediated LDL cholesterol uptake.

Neutrophil elastase 2 (NE, encoded by the ELA2 gene) mutations have been identified as the cause of cyclic and congenital neutropenias, autosomal dominant disorders characterised by a low level of peripheral blood neutrophils and recurring severe bacterial and fungal infections. In cyclic neutropenia, the circulating neutrophil number oscillates from normal to extremely low levels with a 21 day periodicity. In both conditions, neutropenia is associated with a marked elevation in the number of monocytes. As neutrophils and monocytes derived from the same myeloid progenitors, these observations are suggestive of an aberrant switch in cell fate. Normally, NE, expressed in bone marrow progenitors of the granulocytic lineage, is processed through the Golgi apparatus, and the mature NE is stored in azurophil granules until released at sites of inflammation. Its implication in neutropenia suggests that it plays an unexpected role in the regulation of haematopoietic stem cell differentiation. Other candidates are granulocyte colony stimulating factor (GCSF), a growth factor used in treatment of neutropenia, and its receptor (GCSFR), as both are substrates of NE.

Mutations in reelin, an extracellular matrix protein playing a pivotal role in neuronal migration during development, are responsible for lissencephaly, a congenital malformation of the central nervous system (CNS). Reelin binds to the very low density lipoprotein receptor and to the APOE receptor 2 at the cell surface, inducing a signalling cascade to regulate the CNS layer formation. It has also been demonstrated that reelin is a serine protease, of which substrates are extracellular proteins such as laminin and fibronectin. The proteolytic activity of reelin on adhesion molecules of the extracellular matrix and/or neurone receptors may explain how reelin regulates neuronal migration and synaptic plasticity. However, it is not yet known whether the pathogenesis of lissencephaly is dependent or not on its protease activity.

Loss of function mutations in serine proteases of the lysosome lead to two autosomal recessive forms within a group of more than 40 heritable disorders caused by deficiency of lysosomal enzymes. These disorders, grouped under the generic term of lysosomal storage diseases, are characterised by the progressive accumulation of non-metabolised substrates within the lysosome, leading to cellular and tissue damage, organ dysfunction, and early mortality. Deficiency in tripeptidylpeptidase I causes neuronal late infantile ceroid lipofuscinosis (CLN2), a neurodegenerative disorder characterised by seizures, myoclonus, learning disabilities, and ataxia. Mutations in lysosomal carboxypeptidase or cathepsin A are responsible for late infantile and juvenile galactosialidosis, presenting with variable clinical features including skin and neuronal involvements. The resulting complete loss of cathepsin A enzymatic activity leads to a secondary deficiency in 2b-galastosidase. These three enzymes are associated in a large multienzymatic complex in which cathepsin A acts as a protective protein towards the two other enzymes, preventing their degradation.

### Table 1 Serine proteases and their genetic disorders

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein</th>
<th>Disorder</th>
<th>Trans</th>
<th>Function</th>
<th>MIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRSS1</td>
<td>Cathepsin A</td>
<td>Pancreatitis</td>
<td>AD</td>
<td>Digestion</td>
<td>167800</td>
</tr>
<tr>
<td>PRSS7</td>
<td>Enterokinase</td>
<td>Malabsorption</td>
<td>AR</td>
<td>Digestion</td>
<td>226200</td>
</tr>
<tr>
<td>F9</td>
<td>Factor VII</td>
<td>Haemophilia</td>
<td>AR</td>
<td>Procoagulation</td>
<td>227500</td>
</tr>
<tr>
<td>F9</td>
<td>Factor IX</td>
<td>Haemophilia B</td>
<td>X</td>
<td>Procoagulation</td>
<td>306900</td>
</tr>
<tr>
<td>F10</td>
<td>Factor X</td>
<td>Haemophilia</td>
<td>AR</td>
<td>Procoagulation</td>
<td>227400</td>
</tr>
<tr>
<td>F11</td>
<td>Factor XI</td>
<td>Haemophilia C</td>
<td>AR</td>
<td>Procoagulation</td>
<td>264900</td>
</tr>
<tr>
<td>F12</td>
<td>Factor XII</td>
<td>Haemophilia</td>
<td>AR</td>
<td>Procoagulation</td>
<td>234000</td>
</tr>
<tr>
<td>F2</td>
<td>Proheparin</td>
<td>Haemophilia</td>
<td>AR</td>
<td>Procoagulation</td>
<td>176930</td>
</tr>
<tr>
<td>F9</td>
<td>Protease C</td>
<td>Thrombophilia</td>
<td>AD</td>
<td>Procoagulation</td>
<td>176900</td>
</tr>
<tr>
<td>PLAT</td>
<td>Tissue plasminogen</td>
<td>Activator</td>
<td>AR</td>
<td>Anticoagulant</td>
<td>173370</td>
</tr>
<tr>
<td>PLG</td>
<td>Plasminogen</td>
<td>Thrombophilia</td>
<td>AR</td>
<td>Anticoagulant</td>
<td>173350</td>
</tr>
<tr>
<td>C1R</td>
<td>Complement factor I</td>
<td>Complement deficiency</td>
<td>AR</td>
<td>Complement</td>
<td>216950</td>
</tr>
<tr>
<td>C1S</td>
<td>Complement factor I</td>
<td>Complement deficiency</td>
<td>AR</td>
<td>Complement</td>
<td>120580</td>
</tr>
<tr>
<td>C2</td>
<td>Complement factor I</td>
<td>Complement deficiency</td>
<td>AR</td>
<td>Complement</td>
<td>217000</td>
</tr>
<tr>
<td>MASP2</td>
<td>MASP2 protein</td>
<td>Complement deficiency</td>
<td>AR</td>
<td>Complement</td>
<td>605102</td>
</tr>
<tr>
<td>CFI</td>
<td>Complement factor I</td>
<td>Complement deficiency</td>
<td>AR</td>
<td>Complement</td>
<td>217030</td>
</tr>
<tr>
<td>PRSS12</td>
<td>Neutrophastin</td>
<td>Learning disabilities</td>
<td>AR</td>
<td>Remodelling of nervous tissue</td>
<td>249500</td>
</tr>
<tr>
<td>TMPRSS3</td>
<td>Transmembrane protease</td>
<td>Deficiency</td>
<td>AR</td>
<td>Activation of ENaC</td>
<td>605316 DFN810, 601072 DFN8</td>
</tr>
<tr>
<td>PC1</td>
<td>Proheparin convertase</td>
<td>Obesity</td>
<td>AR</td>
<td>Processing of prohormone</td>
<td>600955</td>
</tr>
<tr>
<td>PCSK9</td>
<td>NARC1</td>
<td>Hypercholesterolaemia</td>
<td>AD</td>
<td>Cholesterol level</td>
<td>603776</td>
</tr>
<tr>
<td>ELA2</td>
<td>Neutrophil elastase</td>
<td>Neutropenia</td>
<td>AD</td>
<td>Differentiation myeloid precursors</td>
<td>202700, 162800</td>
</tr>
<tr>
<td>RELN</td>
<td>Reelin</td>
<td>Lissencephaly</td>
<td>AR</td>
<td>Not known</td>
<td>257320</td>
</tr>
<tr>
<td>CLN2</td>
<td>Tripeptidyl-peptidase</td>
<td>Ceroid</td>
<td>AR</td>
<td>Lysosome</td>
<td>204500</td>
</tr>
<tr>
<td>CTSA</td>
<td>Cathepsin A</td>
<td>Galactosialidosis</td>
<td>AR</td>
<td>Lysosome</td>
<td>256540</td>
</tr>
</tbody>
</table>

Trans, mode of transmission; AD, autosomal dominant; AR, autosomal recessive.
GENETIC DISORDERS ASSOCIATED WITH METALLOPROTEASES

Only seven genetic conditions have been associated with mutations in metalloproteases (table 2), even though the metalloproteases are more numerous than serine proteases in the human proteome. Nevertheless, the clinical manifestations and underlying pathological mechanisms are also very diverse. Extracellular matrix remodelling, peptide and protein processing in extracellular or intracellular compartments, and quality control of mitochondrial proteins are altered in these inherited disorders, which affect diverse zinc metalloprotease families (fig 2). Besides these genetic disorders, metalloproteases, especially matrix metalloproteases (MMP), have long been linked with cancer progression. Their activation is increased in almost all human cancers and they have been shown to contribute to tumour invasion through extracellular matrix (ECM) degradation and to tumour angiogenesis.45 MMP have also been associated with neuroinflammation, multiple sclerosis, and rheumatoid arthritis.41 44

Disorders involving matrix metalloproteases

Loss of function mutations in MMP-2 (also known as gelatinase A) have been found in a Saudi family affected by autosomal recessive multicentric osteolysis with arthritis.46 Clinical manifestations include dysmorphic facial features, short stature, marked carpal and tarsal osteopenia, and distal arthropathy progressing towards ankylosis. MMP-2 is expressed in osteoblasts, osteoclasts, and fibroblasts, and was shown to cleave several ECM proteins such as type IV collagen, elastin, fibronectin, and laminin.46 It is postulated that incomplete degradation of the ECM caused by the lack of MMP2 activity could circumvent an appropriate osteoblast bone deposition.

Disorders involving a disintegrin and metalloproteinase with thrombospondin-like motifs (ADAMTS)

Loss of function mutations in ADAMTS2 are responsible for Ehlers-Danlos syndrome type VIIC.47 This autosomal recessive congenital condition is mainly characterised by characteristic facies, extreme skin fragility, and joint laxity. ADAMTS2 is a procollagen N proteinase involved in procollagen processing.48 Collagen maturation involves the cleavage of N and C terminal propeptides to produce mature monomers capable of forming fibrils. The deficiency in ADAMTS2 results in defect in the N terminal processing of procollagen leading to unprocessed precursors. These, once assembled into abnormal “hieroglyphic” collagen fibrils, do not provide the normal tensile strength to skin tissue.

Loss of function mutations in ADAMTS13 lead to thrombotic thrombocytopenic purpura (TTP),50 characterised by intravascular destruction of red blood cell and platelets. ADAMTS13 was identified as the cleaving protease of von Willebrand factor (VWF), the platelet adhesive blood coagulation protein.49 VWF is secreted by vascular endothelial cells and released into plasma as large multimers exhibiting a high affinity for platelets and collagen.50 In normal human plasma, VWF is rapidly cleaved into smaller forms by ADAMTS13. In TTP, the absence of VWF multimer processing results in enhanced platelet aggregation, eventually leading to microvascular thrombosis and haemolysis.

Disorders involving metalloproteases of the neprilysin family

Neprilysin family members are type II transmembrane proteases involved in the extracellular metabolism of biological active peptides.51 Their main role lies in the control of intercellular communication through activation or inactivation of peptidic signals. A mutation in endothelin converting enzyme 1 (ECE-1) was reported in one patient with a form of Hirschspring’s disease (HSCR) associated with cardiac defects and craniofacial abnormalities.52 HSCR is a congenital disorder characterised by a total or partial absence of the nerve cells innervating the intestine. The resulting lack of peristalsis prevents intestinal bolus transit, leading to functional obstruction and megacolon. ECE-1 is involved in the proteolytic maturation of endothelins 1 and 3 (ET1 and ET3), two potent vasoconstrictor peptides important for the development of neural crest derived cells from which the enteric nervous system arises.53 54 The mutation induces a decrease in ECE-1 activities and is therefore expected to reduce the level of mature ET.55 This may result in disturbance of ET mediated intercellular signalling, which is necessary for migration, proliferation, or differentiation of neural crest cells to form enteric nervous ganglia.

Mutations in PHEX (phosphate regulating gene with homologies to endopeptidases located on the X chromosome) endopeptidase cause X linked hypophosphataemia (XLH).56 This disorder is characterised by a renal tubular abnormality resulting in phosphate wasting and defective bone mineralisation, leading to growth retardation and progressive severe skeletal abnormalities. Several studies indicated that a phosphaturic hormone, phosphatonin, may play an important role in the pathophysiological cascade responsible for XLH.57 The inactivating mutations present throughout the entire length of PHEX lead to a failure of phosphatonin clearance from the circulation by abolishing its degradation. Phosphatonin interacts with a renal tubule cell receptor that in turn downregulates a sodium dependent phosphate cotransporter, resulting in an excessive urinary phosphate excretion. Phosphatonin is also thought to be the mineralisation inhibitor shown to accumulate in mutant osteoblasts and responsible for the inhibition of extracellular matrix mineralisation.58

Figure 2 Disorders involving metalloproteases with the organ presenting the prominent phenotype.
Disorders involving other metalloproteases

Mutations in FACE-1 (or ZMPSTE24), a multispanning membrane zinc metalloproteinase, were found in one patient with mandibuloacral dysplasia (MAD). 39 MAD is an autosomal recessive disorder characterised by skeletal abnormalities including hypoplasia of the mandible and clavicles, acro-osteolysis, progeroid appearance, and generalised lipoatrophy. FACE-1 is localised to the ER and the nuclear envelope and is involved in maturation of lamin A, an intermediate filament protein of the nuclear lamina. 40 Processing of prelamin A involves a complex series of post-translational modifications of the C terminus (farnesylation, proteolytic cleavage, and methylation), increasing its hydrophobicity and subsequently facilitating its association with the nuclear lamina. FACE-1 mutations affect prelamin A proteolysis and may compromise its function in nuclear structure, chromatin organisation, or gene regulation.

Mutations in paraplegin, a nuclear encoded mitochondrial metalloprotease cause an autosomal recessive form of hereditary spastic paraplegia (SPG7). 41 SPG7 is characterised by progressive weakness and spasticity of the lower limbs due to selective retrograde degeneration of corticospinal axons. Paraplegin is homologous to yeast mAAA (ATPases associated with a variety of cellular activities) proteases, which associate in a multimeric complex at the mitochondrial inner membrane. This complex exhibits both a chaperone-like activity controlling the assembly of respiratory complexes and a protease activity necessary for the removal of misfolded respiratory chain subunits. 42 SPG7 biopsies showed an accumulation of mitochondriae, signifying a defect in oxidative phosphorylation. 43 A defective respiration seems to be particularly deleterious to long axons, possibly by affecting the energy demanding anterograde and retrograde axonal transport. 44

**GENETIC DISORDERS ASSOCIATED WITH ASPARTYL PROTEASES**

Only one genetic condition is associated with aspartyl proteases and corresponds to the autosomal dominant early onset forms of Alzheimer’s disease caused by mutations in presenilins 1 (PS1) and 2 (PS2) (table 3).

Alzheimer’s disease is characterised by dementia associated with neurone loss, extracellular accumulation of amyloid plaques, and intracellular neurofibrillary tangles of the microtubule associated protein tau. It is the most frequent neurodegenerative disorder, affecting 1% of the population over 65 years of age. Most of the forms are sporadic and age of onset is in the seventies (late onset Alzheimer’s disease). The familial forms are rare and occur before the age of 60 years (early onset Alzheimer’s disease). Mutations in PS1 account for half of the familial cases, and mutations in PS2 for a few percent. Presenilins are heterodimeric membrane proteins generated by endoproteolysis of precursor proteins. 45 They have been shown to correspond to the proteolytic part of γ-secretase, a macromolecular complex involved in intramembrane proteolysis of several transmembrane proteins including amyloid precursor protein (APP) and Notch. 46 47 This processing particularly occurs within the lipidic environment of the membranes and is known as regulated intramembrane proteolysis (RIP). 48 More than 100 mutations have been described in PS1, and eight mutations in PS2 (in both cases, most were missense mutations) have been described in PS1 and PS2, respectively. All result in an accumulation of APP proteolytic fragments, the amyloid-β peptides (Aβ), especially the insoluble neurotoxic Aβ42 form, in the brain. 49 50 Despite intense research, the mechanism by which PS mutations result in increased production of Aβ42 is not fully elucidated, although it is suggestive of an alteration of the selectivity of γ-secretase mediated cleavage.

**GENETIC DISORDERS ASSOCIATED WITH CYSTEINE PROTEASES**

Cysteine proteases are another class of abundant proteases and include the cathepsins, calpains, caspases, deubiquitinating enzymes, and small ubiquitin-like modifier (SUMO) proteases. Two human monogenic conditions have been associated with defects in cathepsin genes, two with caspases, one with a calpain, three for the deubiquitinating enzymes, and none for the SUMO proteases (fig 3, table 4).

Disorders involving cysteine cathepsins

In addition to their well known role in the non-specific degradation of proteins in the lysosomes, the cysteine cathepsins also participate in specialised functions such as prohormone processing, antigen presentation, bone remodelling, spermatogenesis, angiogenesis, apoptosis, and homeostasis of the skin and hair follicles. 51 A recent report provided evidence that they may even play a role within the nucleus, as a cathepsin L isoform was found to proteolytically activate a transcription factor. 52 Increased

---

**Table 2** Metalloproteases and their genetic disorders

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein</th>
<th>Disorder</th>
<th>Trans</th>
<th>Function</th>
<th>MIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADAMTS2</td>
<td>Procollagen N proteinase</td>
<td>Ehlers-Danlos syndrome</td>
<td>AR</td>
<td>Collagen maturation</td>
<td>225410</td>
</tr>
<tr>
<td>ADAMTS13</td>
<td>von Willebrand factor</td>
<td>TIP</td>
<td>AR</td>
<td>VWF cleavage</td>
<td>274150</td>
</tr>
<tr>
<td>MMP2</td>
<td>Matrix metalloprotease</td>
<td>Osteolysis</td>
<td>AR</td>
<td>ECM remodelling</td>
<td>605156</td>
</tr>
<tr>
<td>ECE-1</td>
<td>Endothelin converting</td>
<td>Hirschsprung’s disease</td>
<td>AR</td>
<td>Neural cell crest development</td>
<td>142623</td>
</tr>
<tr>
<td>PHX</td>
<td>PHX</td>
<td>Hypophosphatemia</td>
<td></td>
<td>Clearance of phosphatonin</td>
<td>307800</td>
</tr>
<tr>
<td>ZMPSTE24</td>
<td>FACE1</td>
<td>Mandibuloacral dysplasia</td>
<td>AR</td>
<td>Prelamin A maturation</td>
<td>248370</td>
</tr>
<tr>
<td>SPG7</td>
<td>Paraplegin</td>
<td>Spastic paraplegia</td>
<td>AR</td>
<td>Quality control of mitochondrial proteins</td>
<td>607259</td>
</tr>
</tbody>
</table>

**Table 3** Aspartyl proteases and their genetic disorders

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein</th>
<th>Disorder</th>
<th>Trans</th>
<th>Function</th>
<th>MIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRS1</td>
<td>Presenilin 1</td>
<td>EOAD</td>
<td>AD</td>
<td>generation</td>
<td>AD3 607822</td>
</tr>
<tr>
<td>PRS2</td>
<td>Presenilin 2</td>
<td>EOAD</td>
<td>AD</td>
<td>generation</td>
<td>AD4 600759</td>
</tr>
</tbody>
</table>

EOAD, early onset Alzheimer’s disease; AD, autosomal dominant; AR, autosomal recessive; Trans, mode of transmission.
activity of cathepsins, especially cathepsin B, has been associated with tumour growth, invasion, and metastasis in a number of cancers. Upregulation was also observed in degenerative joint disorders, inflammatory myopathies, and atherosclerosis.

In contrast, the two human genetic pathologies associated with cathepsins correspond to an inactivation.

Loss of function mutations in cathepsin K cause pycnodysostosis, a rare autosomal recessive osteochondrodysplasia. Patients exhibit bone abnormalities such as a short stature, predisposition to bone fractures, cranial deformation, abnormalities in dentition, aplasia of the terminal phalanges, and abnormal density of the entire skeleton. Cathepsin K is strongly and selectively expressed in osteoclasts, specialised cells of the bone that demineralise and digest the bone matrix during the continuous process of bone remodelling. Cathepsin K deficient osteoclasts are no longer capable of degrading the collagen fibres, but demineralisation of bone still occurs. Fibroblast mediated degradation of collagen of soft connective tissues is also affected.

Inactivating mutations in the cathepsin C (dipeptidylpeptidase I) gene result in Papillon-Lefevre or in Haim-Munk syndromes, two rare autosomal recessive disorders. These allelic disorders are characterised by palmoplantar hyperkeratodermy and early onset periodontal destruction leading to tooth loss. Cathepsin C is expressed in skin and gingival epithelia and in osteoclasts and has been implicated in a variety of immune and inflammatory processes. It is implicated in the activation of several lymphocyte and neutrophil serine proteases which participate in cytotoxic apoptosis and in the regulation of cytokine production. The proposed pathophysiological mechanisms leading to the phenotype include a reduced host response against pathogens in the oral cavity, a loss of alveolar bone, an abnormal differentiation affecting the junctional epithelium that normally binds the gingiva to the tooth surface, or a combination of all of those.

Disorders involving calpains

The calpain family comprises intracellular calcium dependent cysteine proteases. The ubiquitously expressed μ and m calpains have been the most widely studied calpains, whereas information on function and regulation of most other members of the calpain family is scarce. Overactivation of ubiquitous calpains following disturbance in calcium homeostasis has been observed in many pathological conditions including heart ischaemia and acute (stroke and trauma) and degenerative neurological disorders (Alzheimer’s and Huntington’s diseases). Calpain 10 appears to be a susceptibility gene for type 2 diabetes in some populations and calpain 9 seems to be a gastric cancer suppressor. However, the only case of monogenic disorder associated with a calpain is recessive limb girdle muscular dystrophy type 2A (LGMD2A) caused by loss of function mutations in the calpain 3 gene.

LGMD2A is characterised by progressive atrophy of the proximal limb muscles, especially those of the posterior compartment, elevated serum creatine kinase, and a necrosis regeneration pattern on muscle biopsies. Calpain 3 is mainly expressed in skeletal muscle, where it can be found either in the nucleus or the cytoplasm, associated with titin, a giant protein of the sarcomere. Among its substrates are the cytoskeletal proteins, filamin C, talin, vinexin, ezrin, and titin. Calpain 3 deficiency alters the survival pathway of NFkB, leading to an increased level of apoptotic myonuclei. It has been proposed that cleavage of the cytoskeletal proteins is part of an adaptive response to mechanical constraints supported by the muscle. Failure to adapt would

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein</th>
<th>Disorder</th>
<th>Transmission</th>
<th>Function</th>
<th>MIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTSK</td>
<td>Cathepsin K</td>
<td>Pycnodysostosis</td>
<td>AR</td>
<td>Bone remodelling</td>
<td>265800</td>
</tr>
<tr>
<td>CTSC</td>
<td>Cathepsin C</td>
<td>Papillon-Lefevre syndrome</td>
<td>AR</td>
<td>Processing of serine proteases</td>
<td>245000</td>
</tr>
<tr>
<td>CASP10</td>
<td>Caspase 10</td>
<td>ALPSI</td>
<td>AD and AR</td>
<td>Immune cell apoptosis</td>
<td>603909</td>
</tr>
<tr>
<td>CASP9</td>
<td>Caspase 8</td>
<td>ALPSI</td>
<td>AR</td>
<td>Immune cell homeostasis</td>
<td>607271</td>
</tr>
<tr>
<td>CAPN9</td>
<td>Calpain 3</td>
<td>LGMD2A</td>
<td>AR</td>
<td>Cytoskeleton remodelling</td>
<td>253600</td>
</tr>
<tr>
<td>UCHL1</td>
<td>Ub C terminal esterase 1</td>
<td>Parkinson’s disease</td>
<td>AD</td>
<td>Degradation of α-synuclein</td>
<td>168600</td>
</tr>
<tr>
<td>CYLD</td>
<td>Cyld protein</td>
<td>Cylindromatosis</td>
<td>AD</td>
<td>Negative regulator of the NFkB pathway</td>
<td>132700</td>
</tr>
<tr>
<td>USP9Y</td>
<td>Ubiquitin specific protease 9</td>
<td>Azoospermy</td>
<td>Y linked</td>
<td>Spermatogenesis</td>
<td>415000</td>
</tr>
</tbody>
</table>

AD, autosomal dominant; AR, autosomal recessive.
Disorders involving caspases

Caspases are known as the proteases of apoptosis, a genetically programmed form of cell death essential for embryonic development, immunity and the pathology of many disorders.48 Apoptotic pathways can be triggered by a number of pathological and physiological stimuli in an extrinsic or intrinsic manner. In the extrinsic pathway, the origin of the death stimulus is external to the cell and involves assembly of an apical death inducing signalling complex (DISC).49 In the intrinsic pathway, the stimulus arises from inside a damaged cell and involves mitochondrial permeabilisation.50 Initial stimuli are followed by a cascade of proteolytic events leading eventually to destruction of architectural components of the cell and DNA fragmentation. Caspases are classified as initiators or effectors depending on their place in the apoptotic cascade. The initiator caspases of the extrinsic pathway are caspases 8 and 10, and that of the intrinsic pathway, caspase 9. An inappropriate level of apoptosis is known to contribute to the pathogenesis of several proliferative or degenerative disorders.51 Excessive apoptosis occurs in traumatic or ischaemic tissue injury, acquired immunodeficiency syndrome, type 1 diabetes mellitus and degenerative neurological disorders such as amyotrophic lateral sclerosis, and Alzheimer’s, Huntington’s, and Parkinson’s diseases. Insufficient apoptosis participates in autoimmune and malignant processes. In particular, somatic mutations in caspases 3, 7, 8, and 10 have been observed in cancer.

Besides these dysregulations, the triggering of which is not well understood in many cases, germline mutations in the two initiator caspases of the extrinsic cell death pathway have been observed in autoimmune lymphoproliferative syndrome type II (ALPSII). ALPSII is characterised by an aberrant proliferation of lymphocytes, resulting in adenoma
thes and splenomegaly, and is associated with autoimmune and inflammatory manifestations. Two caspase 10 mutations (one recessive, one dominant) have been found in ALPSII families,52 and the study showed that the dominant mutation acts in a dominant negative manner, probably at the DISC level, to impair the death receptor induced apoptosis in lymphocytes and dendritic cells, whereas the recessive mutation has a less severe apoptotic defect. Homozygous individuals for a caspase 8 mutation, presenting ALPS symptoms together with immunodeficiency, have been identified in one family.53 In addition to inducing defective lymphocyte apoptosis by a decrease in caspase 8 recruitment to the DISC, the mutation impairs the activation of T and B lymphocytes and of natural killer cells, as demonstrated by the absence of activation markers at the cell surface and defects in cytokine and immunoglobulin production.

Disorders involving deubiquitinating enzymes

Ubiquitination is a reversible post-translational modification of proteins, which controls many cellular functions. An ubiquitin chain consisting of at least four ubiquitins constitutes a signal of degradation by the proteasome.54 Other modifications are implicated in regulatory mechanisms such as control of receptor internalisation, intracellular trafficking, DNA repair, or transcriptional activity. De-ubiquitinating enzymes (DUB) are cysteine proteases of the ubiquitin pathway that specifically cleave ubiquitin from ubiquitin precursors and ubiquitin protein conjugates, ensuring correct equilibrium of these processes with ubiquitin recycling.55 To date, the DUBs have been divided into the ubiquitin C terminal hydrolase (UCH) and the ubiquitin specific processing protease (UBP or USP) families, on the basis of conserved sequence motifs.56 In the human genome, four and 63 distinctive genes encoding UCH and UBP respectively, have been identified. An ever increasing number of mutations are being found in DUB.

A mutation in ubiquitin carboxyterminal esterase L1 (UCHL1) was identified in a family with a dominant form of Parkinson’s disease,57 characterised by a selective loss of dopaminergic neurones and the presence of Lewy bodies, inclusions mainly composed of α-synuclein, a presynaptic protein implicated in neuronal plasticity or vesicle transport.58 In addition to its hydrolase activity, UCHL1 has also been shown to possess ubiquitin ligase activity.59 The mutated UCHL1 has a reduced catalytic activity that may affect the degradation of α-synuclein leading to its accumulation and aggregation.60 Interestingly, a polymorphic variant of UCHL1 with reduced ligase activity but comparable hydrolase activity with the wildtype enzyme has been associated with a decreased susceptibility for Parkinson’s disease.61

Truncating mutations in the tumor suppressor CYLD1 cause cylindromatosis, an autosomal dominant predisposition to benign skin tumours.62–64 CYLD1 has sequence homology to UCH and has been shown to negatively regulate the NFKb pathway through deubiquitination of TNFα receptor associated factors (TRAF), major mediators of the TNF signalling.65–67 CYLD dysfunction results in excessive ubiquitination of TRAF, an activation signal in this pathway.100 The consequences are activation of IkB kinase (IKK), subsequent degradation of IkBα, the inhibitor of NFKb, and ultimately excessive NFкB activation. This activation triggers cell transformation through increased resistance to apopto
sic.68

Loss of function mutations in ubiquitin specific protease 9 (USP9Y; also named DFFRY) is associated with azoospermia, the absence of sperm production.101 USP9Y is a Y chromosome gene that possesses a homologue on the X chromosome (DFFRX), both presenting homology with Drosophila developmental gene fat facets (faf),102 important for the normal development of embryo and eye, preventing the proteasomal degradation of specific proteins.63 By similarity, efficient progression of spermatogenesis might require the stabilisation of particular proteins.

GENETIC DISORDERS OF OTHER COMPONENTS OF PROTEOLYTIC SYSTEMS

This review focuses on proteases. Nevertheless, it is worth mentioning briefly instances in which mutations in other components of proteolytic systems contribute directly to disease pathogenesis. For example, several human disorders have been associated with mutations in protease inhibitors such as serpins, cystatins, and tissue inhibitors of metalloproteinases, which are inhibitors of serine, cysteine and metalloproteases, respectively.104–106 Defective proteasome-dependent proteolysis is also encountered in cases of inefficient substrate recognition caused by mutations in E3 ubiquitin ligases or in substrates themselves, often leading to a toxic accumulation of substrates within the cells. E3 ubiquitin ligases recognise and catalyse the ubiquitin conjugation to specific substrates for degradation by the proteasome or modulation of protein activities.55–56 Five human disorders have been associated to date with mutations in such enzymes: limb girdle muscular dystrophy type 2H, Angelman syndrome, Lafora disease, autoimmune polyendocrinopathy candidiasis ectodermal dysplasia, a form of Fanconi anaemia, and an autosomal recessive form of juvenile Parkinson’s disease.107–108 It is interesting to note that most of the disorders associated by E3 ubiquitin ligases are transmitted in a recessive manner, whereas those associated with DUB are transmitted in a dominant fashion. Alterations of the structure of a substrate can also modify its degradation by the ubiquitin/proteasome system. In
particular, mutation can result in misfolding of the protein, rendering it susceptible to aggregation and preventing it from being efficiently recognised and degraded by the ubiquitin/proteasome machinery. This mechanism underlies the pathogenesis of several major human neurodegenerative disorders: a dominant form of Parkinson’s disease with accumulation of α-synuclein, Alzheimer’s disease with accumulation of amyloid precursor protein, and Huntington’s disease and spinocerebellar ataxias associated with polyglutamine expansions. Furthermore, there is supporting evidence that the protein aggregates can compromise the ubiquitin/proteasome function by secondary sequestration of its components.

CONCLUSION AND PERSPECTIVES

To date, 42 inherited disorders have been identified in a total of approximately 500 proteases. As this accounts for less than 10%, it is likely that they represent only a fraction of all existing disorders. It should be noted that the majority of these disorders are of recessive inheritance with loss of function mutations, indicating absolute requirement of these proteolytic activities for normal cellular function. All the examples presented in this review illustrate the role of proteolysis in processes as diverse as digestion, coagulation, regulation of the immune system, regulation of channel activity, peptide and hormone metabolism, gene regulation, clearance and quality control of proteins and tissue development, differentiation, and plasticity. This impressive list highlights the importance for a cell to have tools at its disposal performing irreversible modifications in order to orientate signalling.

Identification of a causative gene helps to understand its physiological function by providing a unique opportunity to examine the consequences of the perturbation of a specific protein in human beings (depending on the availability of sample materials). This is pivotal in the cases where the corresponding mouse model does not manifest the phenotype of the human disorder or when a model is not even feasible, as in case of the absence of orthogonal genes. Caspase 8 is a particularly striking example in this aspect. Caspase 8 disruption is lethal in mice, whereas in humans it presents itself as a lymphoproliferative and immunodeficiency syndrome. As caspase 10, a close caspase 8 parologue, does not exist in the mouse, this divergence was inferred to arise by virtue of partial functional redundancy between the two human proteins. In addition, the phenotype of caspase 8 deficiency in humans includes a default in immune cell maturation, indicative of a non-apoptotic function of caspase 8 that may not have been uncovered otherwise.

Another remarkable example of an unexpected function unravelled by the detection of causative mutation is neutrophil elastase. While this enzyme is well known for its destructive effects on tissues at the site of inflammation, its implication in neutropenia has revealed an unforeseen role in cell fate determination.

Identification of the implicated protein and elucidation of the molecular mechanisms leading to the disorder will help to develop new therapeutic agents capable of reversing abnormal phenotypes. Strategies should be adapted depending on whether there is an overactivation or an inactivation of the implicated protease. An overactivation of a protease could be theoretically counterbalanced by any means to suppress this activation: specific inhibitor or small RNA. Loss of proteolytic activity could be compensated by replacing the defective enzyme with its normal counterpart either as protein or as coding sequence or by intervention downstream of its physiological action. Many of these strategies are promising and should give rise to new efficient therapeutic agents for these disorders in the near future. They may also prove useful in pathological situations such as cancer or inflammation, in which the proteolytic systems are deregulated as secondary pathological consequences.

ACKNOWLEDGEMENTS

I would like to acknowledge Dr D Bechet for the initial idea of this manuscript. I thank Drs N Daniele, S Lupton, M Bartoli, A Bernot, and O Donos for critical reading of the manuscript and helpful suggestions. This work was supported by the Association Française contre les Myopathies.

Competing interests: none declared

REFERENCES


www.jmedgenet.com


Hong SE, Shograt YY, Huang DT, Shahwan SA, Grant PE, Hourihan JO, Martin ND, Walsh CA. Autosomal recessive lissencephaly with cerebellar atrophy (LSA3) is associated with neutropenia in dogs and humans disrupt intracellular transport of neutrophil elastase. Nat Genet 2003;35:90–6.


The genetic and molecular bases of monogenic disorders affecting proteolytic systems

539


Normal BRCA1 DNA-binding activity in BRCA1-mutated tumors is associated with increased expression of cell cycle regulators and low level of apoptosis. Cancer Res 2003;63:2775–82.


The genetic and molecular bases of monogenic disorders affecting proteolytic systems

I Richard

doi: 10.1136/jmg.2004.028118

Updated information and services can be found at:
http://jmg.bmj.com/content/42/7/529

These include:

References
This article cites 120 articles, 29 of which you can access for free at:
http://jmg.bmj.com/content/42/7/529#BIBL

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/