# II. LITERATURE REVIEW

## A. Dengue virus

### 1. Structure and composition of the virion

Dengue viruses are members of the Family *Flaviviridae*. There are four distinct serotypes of dengue virus, namely, dengue virus serotype 1 (DEN-1), dengue virus serotype 2 (DEN-2), dengue virus serotype 3 (DEN-3), and dengue virus serotype 4 (DEN-4). Dengue virus virions consist of a single-stranded linear RNA molecule, which is capped at the 5'end, but lacks a poly(A)$^+$ tract at the 3'end (Henchal and Putnak, 1990). The RNA genome is about 11 kilobases in length, surrounded by an icosahedral nucleocapsid of about 30 nm in diameter. This is covered by a lipid envelope of about 10 nm in thickness. The complete virion is about 50 nm in diameter (Henchal and Putnak, 1990).

### 2. Genome structure

Complete nucleotide sequences of the genomic RNA of all four dengue virus serotypes have been described. The order of the coding sequences is 5' C - prM(M) - E - NS1 - NS2A - NS2B - NS3 - NS4A - NS4B - NS5 3'. Sequences of one-fourth of the genome at 5'end encoded the structural proteins, whereas the non-structural proteins are encoded in three-fourth of the genome at 3'end. The genomic RNA has a type I cap at its 5'end ($m^7$GpppAmp) followed by the conserved dinucleotide sequence AG and ends with the conserved dinucleotide CU (Rice, 1996).

The dengue virus genome contains a single long open reading frame of more than 10,000 nucleotides. The protein derived from translation of the viral genomic RNA is a single, long, precursor polyprotein that is successively cleaved by both host and viral-encoded proteases to produce structural and non-structural proteins (Rice, 1996). Flanking the open reading frame are the 5' and 3'nontranslated regions, both

of which contain conserved RNA sequences. At the 3'end of the genomic RNA, there is a secondary structure of about 90 bases, which appears to conserve both information and stability (Brinton *et al.*, 1986). There is a short conserved RNA sequence, CS1, which is about 26 nucleotides in length, located on the 5'side of the secondary structure 3'end. Part of this CS1 is complementary to a conserved sequence that is located in the region encoding the capsid protein (5'CS). This complementary region may function in the cyclization of the viral genome; this, in turn, may be important in the regulation of translation, replication or packaging. On the 5'side of CS1, there is an additional conserved sequence, CS2, which is about 24 nucleotides in length. This sequence is duplicated in all four serotypes of dengue virus, but the significance of this CS2 is not yet known (Rice, 1996).

### 3. Viral proteins

Dengue RNA consists of a long open reading frame and encodes 10 viral proteins, three of which are structural (C, prM and E) and seven non-structural (NS) proteins. Result translation of this genome is a single polyprotein cleaved by enzymes that derived from the cell and virus, which separate the polyprotein into individual proteins (Henchal and Putnak, 1990).

#### 3.1. Viral structural proteins

The mature virion contains three structural proteins: C protein, the nucleocapsid or core protein; M protein, a membrane-associated protein; and E protein, the envelope protein. The immature virions are mainly located intracellularly contain prM protein, which is an M protein precursor.

The C protein is the first viral polypeptide synthesized during translation and it has a molecular weight of 9 to 12 kDa and contains 112 to 127 amino acids. This protein has a highly positive charge because of large numbers of lysine and arginine

residues (25%). This highly basic character is believed to be associated with the negative charged viral RNA molecule that forms the nucleocapsid (Hahn *et al.*, 1988). The C protein lacks an *N*-terminal, hydrophobic signal sequence, which suggests that its synthesis is on the non-membrane-bound ribosome. A hydrophobic stretch of amino acid at the C protein *C*-terminus probably acts as the transmembrane secretion signal for prM. This hydrophobic domain may serve to transiently anchor the C proteins to a membrane at the replication site after cleavage, probably by a host cell "signalase" at the *N*-terminus of the prM protein (Monson, 1989).

The prM protein is the glycosylated precursor of the structural protein M (18.1 to 19.1 kDa) and is contained in intracellular immature virions that form a heterodimer with the E protein (Wangler and Wangler, 1989). The E-prM interaction is believed to protect the E protein from irreversible inactivation during transport to the cell surface in acidic post-Golgi vesicles (Konishi and Mason, 1993). The prM protein contains six cysteine residues that form three disulfide bridges. However, during viral release, this protein is generally cleaved at acidic post-Golgi vesicles to form the M and the *N*-terminal pr segment. This cleavage occurs shortly before or simultaneously with virion release. The structural protein M (7 to 9 kDa) is located in the *C*-terminal portion of prM and is present in mature virions (Rice, 1996; and Henchal and Putnak, 1990). The cleavage is mediated by the host enzyme furin or other enzymes of similar specificity (Randolph *et al.*, 1990; Stadler *et al.*, 1997).

This cleavage, which may occur in the acidic post-Golgi vesicle, appears to pre-precede the virus release from the cell. The formation of M from prM appears to be the crucial, terminal event in virion morphogenesis. It results in a large increase in viral infectivity and reorganization of the viral surface structure that is composed of E-prM heterodimers in immature virions (Wangler and Wangler, 1989; Allison *et al.*, 1995). The association of prM with the E protein in immature virions appears to make the latter protein resistant to the low pH, which induces the conformational change

required for fusion. This association has been shown to stabilize a pH sensitive epitope on the E protein and prevents viral aggregation at an acidic pH (Heinz *et al.*, 1994). The prM protein cleavage by furin in vitro shows a 100-fold increase in specific infectivity, the acquisition of membrane fusion and hemagglutination activity. Also, the ability of the envelope proteins can undergo low-pH induced structural rearrangement that is characteristic of mature virions (Stadler *et al.*, 1997). Cleavage of the prM can be blocked by the treatment of infected cells with ammonium chloride or bafilomycin A1, resulting in the release of virus particles containing the uncleaved prM precursor protein. These immature virions are less infectious than mature ones (Wangler and Wangler, 1989; Randolph *et al.*, 1990; Heinz *et al.*, 1994) and do not induce cell-cell fusion at an acidic pH (Guirakhoo *et al.*, 1991; Guirakhoo *et al.*, 1993).

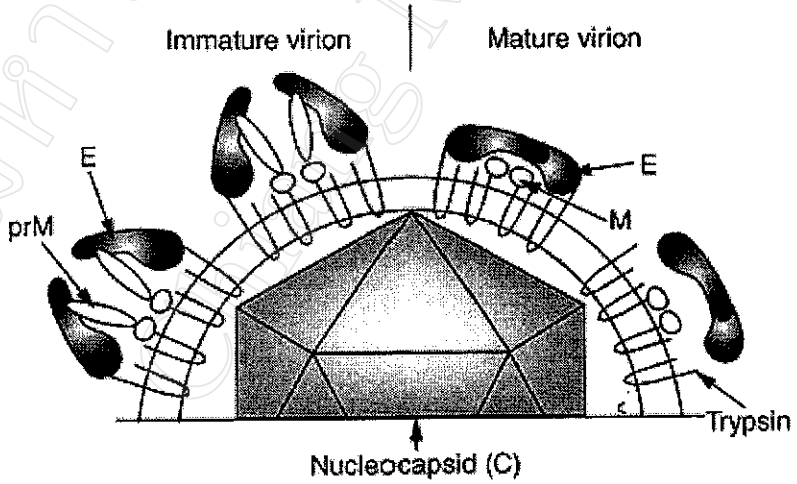



**Figure 1.** Schematic diagram of the composition of immature and mature flaviviruses (Figure reproduced from Heinz and Allison, 2001.)

The E glycoprotein (55 to 60 kDa) is the major protein component of the virion surface. The E protein plays a role in several important biological activities including receptor binding, hemagglutination of goose erythrocytes, fusion activity in acid pH endosomes virus assembly and induction of protective immunity (Heinz *et al.*, 1994; Raviprakash *et al.*, 2000). Purified E glycoprotein can elicit neutralizing antibodies against dengue serotype 2 virus and partially protect mice from fatal infection (Feighny *et al.*, 1992). The E protein is a transmembrane protein with a *C*-terminal hydrophobic region inserted in the lipid bilayer of the envelope. This hydrophobic part is required for cleavage at the E/NS1 junction by the host protease (Falgout *et al.*, 1989).

The E glycoprotein is found to be a homodimer or homotrimer on the surface of mature virions and it has two potential *N*-linked glycosylation sites, Asn-67 and Asn-153. These glycosylation sites may effect virus-mediated membrane fusion and neurovirulence (Guirakhoo *et al.*, 1993).

The structure of the flavivirus E proteins has been determined to 2-Å resolution by X-ray crystallography of a soluble dimeric fragment (ectodomain), isolated by limited trypsin digestion from purified Tick-borne encephalitis virions (Rey *et al.*, 1995). Although the Tick-borne encephalitis virus is the only one flavivirus in which the E protein structure has been determined so far, it can be assumed, based on sequence homology and the conservation of all six disulfide bridges (Henchal and Putnak, 1990), that E proteins of approximately 70 different flaviviruses share the same basic architecture. The soluble E protein dimer of flaviviruses folds into three distinct domains, namely, a central $\beta$-barrel (domain I), an elongated dimerization region (domain II) and a *C*-terminal, immunoglobulin-like module (domain III). In all three domains, $\beta$-strand secondary structure is predominate. The domain I contains about 120 residues that derive from three non-contiguous segments (residues 1-51, 137-189, 285-302) of the whole protein. The

domain II structure (residues 52-136, 190-284) contains a highly conserved sequence (residues 98-113) and is stabilized by five disulfide bridges. This domain is important for the fusogenic activity of the viral envelope with host cell membranes (Roehrig *et al.*, 1990). The domain III structure (residues 303-395) requires a disulfide bond between Cys-11 and Cys-12. Domain III is likely to be involved in receptor binding. It was demonstrated that monoclonal antibodies to epitopes, which located in domain II and III, were able to block viral hemagglutination and neutralize virus infectivity. Domains I, II and III constitute the head of the E glycoprotein. A 100-amino acid stalk (approximate residues 400-500), which includes a 50 amino acid hydrophobic tail (residues 450-500), anchors the molecule in the virion envelope. The E protein is a homodimer with its subunits arranged in head-to-tail orientation and lies flat, parallel to the viral membrane. The upper surfaces of the extodomains carry the antigenic determinants, which are recognized by neutralizing antibodies, as well as a single *N*-linked carbohydrate side chain attached to each of the monomers. Each of the monomers is composed of three distinct domains (I, II and III) that have predominantly $\beta$-strand secondary structure and are connected by potential hinge regions. The loop at the tip of domain II contains the fusion peptide, which is constrained by multiple interactions including a disulfide bridge, and its top surface is covered by the carbohydrate side chain of other monomers (Roehrig *et al.*, 1989; Rey *et al.*, 1995; Stuart and Gouet, 1995; Allison *et al.*, 2001; Ferlenghi *et al.*, 2001).

Mapping of the antigenic epitopes by a serological method indicates that these domains are three clusters of antigenic epitopes, which designate A, B and C, on the E protein (Mandl *et al.*, 1989). These antigenic clusters correspond well with the domains II, III and I, respectively, of the three-dimensional structures. Flavivirus cross-reactive epitopes are found in the domain A, whereas, type or subtype specific epitopes are found in the other two domains (Rey *et al.*, 1995).

### 3.2. Viral non-structural proteins

Seven viral non-structural (NS) proteins encoded in 3'terminus of polyprotein follow the E protein. The order of the non-structural proteins coding from the 5'end to 3'end is 5' NS1 - NS2A- NS2B - NS3 - NS4A - NS4B - NS5 3'.

NS1, the first non-structural protein, is a 42 to 52 kDa. NS1 includes 12 strictly conserved cysteine residues, one to three *N*-linked glycosylation sites, and regions of high sequence conservation. The function of NS1 in viral replication is unknown.

The NS2 consists of two proteins, NS2a and NS2b. NS2a (18 to 22 kDa) is required for proper proteolytic processing of the *C*-terminus of NS1 (Falgout and Markoff, 1995). NS2b (13 to 15 kDa) is involved in the protease function of the NS2b- NS3 complex (Preugschat *et al.*, 1990).

The NS3, a hydrophilic protein (Molecular weight ~68 to 70 kDa), is highly conserved among flaviviruses and is believed to be an enzymatic component of the RNA replication machinery. NS3 is a large trifunctional protein, containing protease, helicase, and RNA triphosphatase activities. The *N*-terminal region of NS3 contains conserved amino acids shared by trypsin-like serine protease, whereas, the *C*-terminal domain is homologous to helicases (Bazan and Fletterick, 1989; Gorbalenya *et al.*, 1998). RNA-stimulated NTPase activity has been demonstrated in the purified NS3 (Wangler and Wangler, 1991). The presence of proteinase and helicase domains within a single viral protein is a common structural motif found in alphavirus, pestivirus, potyvirus, and coronavirus proteins as well as flaviviruses. The primary function of a viral proteinase is the regulating posttranslational production of individual gene products from a polyprotein precursor. The viral proteinase can also produce processing intermediates, which may themselves be functional components of the viral life cycle (Preugschat *et al.*, 1990).

The NS4a and NS4b are hydrophobic proteins, which have molecular weights of 16 kDa and 27 kDa, respectively. They might be RNA replication complex cofactors along with the putative viral RNA-dependent RNA polymerase, NS5.

The NS5 protein is the last protein encoded in the long open reading frame, the largest (103 to 104 kDa) and most highly conserved flavivirus protein. NS5 is a basic protein, lacking of long hydrophobic stretches, and is believed to be the flavivirus RNA-dependent RNA polymerase. The *N*-terminal domain of NS5 (between residues 60 and 145) is homologous to a region of methyltransferases implicated in S-adenosylmethionine binding. It has been suggested that this domain may be involved in methylation of the 5'cap structure (Rice, 1996). The NS5 protein is at least a bifunctional protein, possessing both methyltransferase and RNA polymerase activities.

**4. Synthesis and processing of viral proteins**

The viral structural and non-structural proteins derive from a large precursor polyprotein encoded by a long open reading frame. This precursor protein is not usually seen in the infected cells. Translation begins at the first AUG codon of RNA genome, and then individual processing of the precursor peptide occurs. The proteases associated with polyprotein processing appear to be host and virus encoded. The host "signalase" in the lumen of the endoplasmic reticulum is cleaved at the *N*-termini of prM, E, NS1, and NS4b. These cleavage sites are hydrophobic signal-like sequences, which are required for proteolytic processing of precursor proteins (Markoff, 1989).

The cleavage at the *N*-termini of NS2b, NS3, NS4a and NS5 are mediated in the cytosol by the NS2b-NS3 protease complex.

The prM cleavage site sequence, -Arg(P4)-Xaa(P3)-Arg/Lys(P2)-Arg(P1)↓- (where Xaa is variable and ↓ represents the cleavage site where the peptide is

hydrolyzed), which completely fits the consensus sequence for cleavage by furin, is similar to that used in the maturation cleavage of numerous viral glycoproteins (Miranda *et al.*, 1996; Stadler *et al.*, 1997; Sabbarao *et al.*, 1998; Volchkov *et al.*, 1998).

## B. Subtilisin-like proprotein convertases

### 1. Members of subtilisin-like proprotein convertase family in mammalian

The processing of precursor proteins via limited proteolysis is an important and widely used cellular mechanism for the generation of biologically active proteins and peptides in appropriate cellular compartments. The major endoproteolytic processing enzymes of the secretory pathways are the homologous endoproteases with bacterial subtilisin or the yeast processing protease, Kexin (Kex2p) (Nakayana, 1997; Steiner, 1998; Zhou *et al.*, 1999), which is encoded by the *fur* gene on chromosome 15. Subsequently, seven mammalian Kexin-homologue enzymes were identified. Although each enzyme had been independently named by it discoverers, a simplified nomenclature for the group of mammalian processing proteases has been proposed (Chan *et al.*, 1992; Bergeron *et al.*, 2000), using the term subtilisin-like proprotein convertases (SPCs):

SPC1 [Furin or pair amino acid convertase enzyme (PACE)],

SPC2 [Prohormone convertase 2 (PC2)],

SPC3 [Prohormone convertase 1 (PC1) or Prohormone convertase 3 (PC3)],

SPC4 [Pair amino acid convertase enzyme 4 (PACE4)],

SPC5 [Prohormone convertase 4 (PC4)],

SPC6 [Prohormone convertase 5 (PC5) or Prohormone convertase 6A (PC6A) and SPC7 [Lymphoma prohormone convertase (LPC) or Prohormone convertase 7 (PC7) or Prohormone convertase 8 (PC8)].

**2. Structure and function of subtilisin-like proprotein convertases**

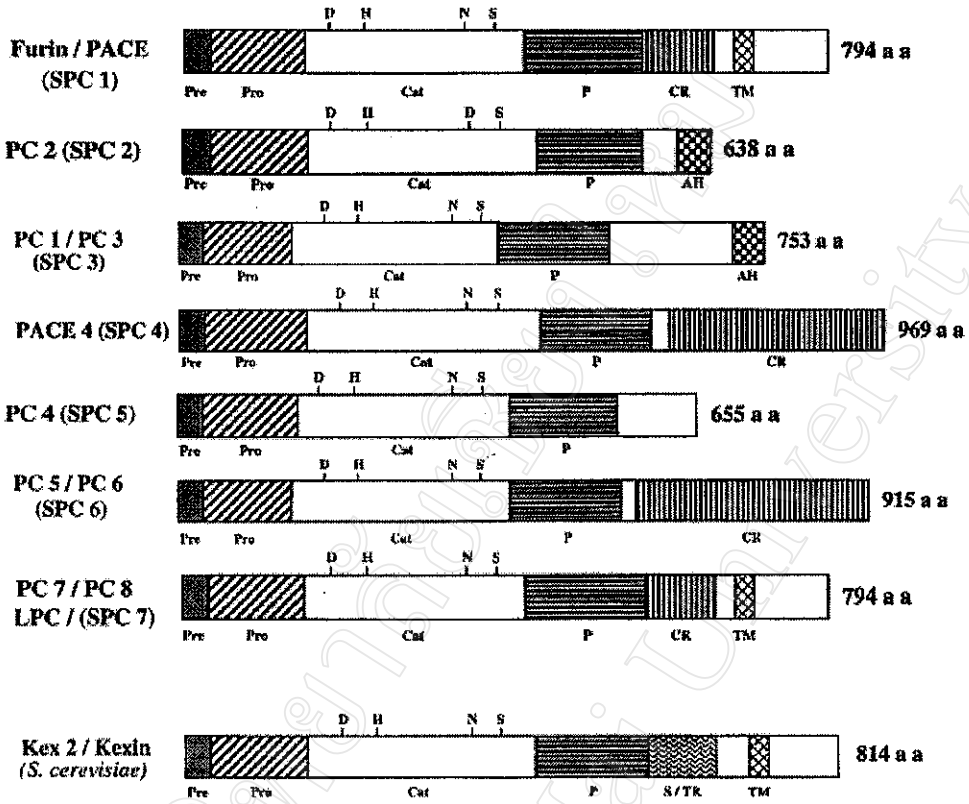
These subtilisin-like proprotein convertases all have a characteristic amino-terminal propeptide (Pro). The Pro is followed by a well conserved, but modified (from that of subtilisin); catalytic module (Cat) and a conserved down-stream domain of 150 amino acids called the "P domain" or "homo B-domain". A three-dimensional structure has not yet been obtained for any of these enzymes, although their catalytic modules have been modeled on the basis of the known structure of subtilisin (Lipkind *et al.*, 1998). The catalytic domain contains the active site of the enzyme, with the typical catalytic triad of subtilisin-related serine proteases, including the asparagine, histidine and serine active site residues. The role of the P domain appears to be regulatory and it may influence the marked calcium dependency and increased acidic pH optima of some of the subtilisin-like proprotein convertases. In addition, the P domain also appears to stabilize the catalytic domain structurally. It may do this by helping to balance the surface charge asymmetry in the substrate-binding region of the catalytic domain. This is caused by the characteristic multibasic residue specificity of these enzymes. The roles of the substrate recognition of subtilisin-like proprotein convertases have been identified by mutational analysis confirming that negative charged residues of a catalytic pocket interact with positive charged residues of the substrate (Creemers *et al.*, 1993). The consensus substrate sequence is –Arg(P4)-Xaa(P3)-Lys/Arg(P2)-Arg(P1)↓-Xaa(P1')-, where Xaa can be almost any amino acid and ↓ represents the cleavage site where the peptide bone is hydrolyzed.

SPC1 (furin/PACE) is a secretory pathway endoprotease that catalyses the maturation of a strikingly diverse group of proprotein substrates, ranging from growth factors and receptors to pathogen proteins in multiple compartments within the *trans*-Golgi network/endosomal system. Human furin is initially synthesized as 100 kDa core-glycosylated pro-furin, which is converted into 94 kDa forms by a cleavage of

propeptide at the Arg-Ala-Lys-Arg- site, at residue 104-107 (Denault and Luduc, 1995; Nakayama, 1997).

SPC2 (PC2) is more complex in transport and activation. This convertase, which is unique, requires the acidic conditions of a late post-Golgi compartment for activation. In the ER, proSPC2 interacts with 7B2: a 27-kDa neuroendocrine secretory protein that is coexpressed with SPC2 in many neuroendocrine tissues. The 7B2 is required for the production of active SPC2, which might function as a chaperone protein to assist in proSPC2 folding. The 7B2 is cleaved at a polybasic site toward the *C*-terminus, similar to the cleavage of furin or related TGN protease, which results in the release of an *N*-terminal 21-kDa form and inhibitory *C*-terminal fragment (Steiner, 1998; Bergeron *et al.*, 2000).

The other SPC: SPC3 (PC1/PC3), SPC4 (PACE4), SPC5 (PC5/PC6), SPC7 (LPC/PC7/PC8), have sequence specificity similar to that of SPC1 (furin). The mammalian subtilisin-like proprotein convertases can be classified into three groups based on their tissue distribution. SPC1 (furin), SPC4 (PACE4), SPC6 (PC5/PC6) and SPC7 (LPC/PC7/ PC8) are expressed in a broad range of tissues and cell lines. In contrast, the expression of SPC2 (PC2) and SPC3 (PC1/PC3) is limited to neuroendocrine tissues, such as pancreatic islets, pituitary, adrenal medulla and many brain areas. The expression of SPC5 (PC4) is highly restricted to testicular spermatogenic cells. Within cells, SPC1 (furin) and SPC7 (LPC/PC7/PC8), both of which have a transmembrane domain, are localized in the *trans*-Golgi network (TGN). Another SPC with a transmembrane domain, SPC6B (PC5/PC6B), is also localized in the Golgi area, although it appears not to concentrate in the *trans*-Golgi network. The neuroendocrine-specific convertases, SPC2 (PC2) and SPC3 (PC1/PC3), are mainly localized in secretory granules. SPC6A (PC5/PC6A) has been reported to be localized to secretory granules (Nakayama, 1997; Seidah and Chretien, 1997).




Letters above the open boxes represent the catalytic triad (D, H, S) and the oxyanion residue (N; D only in PC2). Other abbreviations are as follows: Pre, signal sequence; Pro, proregion; Cat, catalytic domain; P, P or homo B domain; AH, predicted amphipathic helix; CR, cysteine-rich region; TM, transmembrane domain; S/TR, serine/threonine-rich region.

**Figure 2.** Members of the mammalian family of subtilisin-like proprotein convertase compare with Kex2 (Kexin). [Figure modified from Steiner (1998)].

The autoactivation of SPC1 (furin) serves as a model for the other subtilisin-like proprotein convertases, with the exception of SPC2 (PC2). Intramolecular cleavage of the propeptide (Pro) allows furin to exit the endoplasmic reticulum. However, the propeptide remains attached noncovalently until the cleaved inactive proenzyme reaches the *trans*-Golgi network, where an increased acidic (pH~6.5) and calcium-enriched environment facilitates dissociation from the propeptide. A second cleavage within the propeptide then precludes further inhibitory interactions, which result in full activation. A similar mechanism of activation has been demonstrated for SPC3 (PC1/PC3), SPC5 (PC4) and SPC7 (LPC/PC7/PC8). Propeptide of SPC4 (PACE4) is autocleaved slowly, but it can probably also occur prior to its exit from the endoplasmic reticulum (Zhou *et al.*, 1999).

In the C6/36 mosquito cell-line, the pro-vitellogenin convertase deduced amino acid sequence has a high similarity to a domain structure characteristic of subtilisin-like convertases. Chen and Raikhel (1996) identified and characterize a 115-kDa pro-vitellogenins (pro-Vg) processing enzyme, pro-vitellogenins convertase (VC), from a vitellogenic female fat body cDNA library of *Aedes aegypti* mosquito. The amino acid sequence of vitellogenins convertase reveals a high structural similality to furin-like convertase. It has a typical structure of furin-like convertases which comprised of pre-propeptide, catalytic, cysteine-rich, *C*-terminal transmembrane, and cytoplasmic domains and recognizes the same cleavage sites that contain paired basic amino acid motif, Arg(P4)-Xnn(P3)-Lys/Arg(P2)-Arg(P1)↓-.

### 3. Biochemical and enzymatic feature of furin

Consistent with furin as the first and so forth the best-characterized enzyme of subtilisin-like proprotein convertase, furin's enzymatic activity and specificity has been evaluated in vitro and in vivo. *In vitro* assays, relying on the use of either the shed form of furin and/or the soluble furin measure, the cleavage of various substrates

such as the fluorogenic tetrapeptides or use of internally quenched fluorogenic peptidyl substrates (Krysan, 1999). Krysan found that furin could cleave such substrates with kinetics comparable to those observed with extended peptides and physiological substrates. With the best of these hexapeptidyl methylcoumarinamides, furin displayed $k_{cat}/K_M$ values greater than $10^6$ $M^{-1}s^{-1}$. Furin exhibited striking substrate inhibition with hexapeptide but not with tetrapeptide substrates, an observation of significance to the evaluation of peptide-based furin inhibitors. Quantitative comparison of furin and Kexin recognition at P1, P2, and P4 demonstrates that whereas interactions at P1 make comparable contributions to catalysis by the two enzymes, furin and Kexin exhibited a 10-fold lesser dependence on P2 recognition. However, a 10-100 fold greater dependence on P4 recognition was observed. Furin has recently been shown to exhibit P6 recognition and Krysan found that this interaction contributes 1.4 kcal/mol toward catalysis independence of the nature of the P4 residue. Krysan has also shown that those favorable residues at P2 and P6 compensate for fewer optimal residues at either P1 or P4. The quantitative analysis of furin specificity sharply distinguishes the nature of substrate recognition by the processing and degradative members of subtilisin-related proteases.

**4. Target of subtilisin-like proprotein convertase in the surface proteins of other viruses**

As listed in Table 1, the cleavage sites of these flavivirus prM precursors and other viral coated glycoproteins fully fit the –Arg(P4)-Xaa(P3)-Lys/Arg(P2)-Arg(P1)-consensus sequence. These viral precursors were delineated the following sequence rules that govern the cleavage by subtilisin-like proprotein convertases, as follows: -

- An Arg residue is essential at the P1 position.

- In addition to the P1 Arg, basic residues at the P2 and P4 are required for efficient cleavage.

- At P1' position, an amino acid with a hydrophobic aliphatic side chain is not suitable.

**Table 1.** Sequences around the cleavage site of precursor proteins

| Precursors | GenBank no. | Cleavage site | | | | | | | References |
|---|---|---|---|---|---|---|---|---|---|
| **Flaviviruses prM proteins** | | **... P6** | **P5** | **P4** | **P3** | **P2** | **P1 ↓** | **P1' ...** | |
| Deangue virus group | | | | | | | | | |
| Dengue virus type 1 (West Pacific) | GI: 1854036 | ... H | R | R | D | K | R | $S_{206}$ ... | Puri *et al.*, 1997. |
| Dengue virus type 2 (16681) | GI: 2155257 | ... H | R | R | E | K | R | $S_{206}$ ... | Kinney *et al.*, 1997. |
| Dengue virus type 3 ( H87) | GI: 323468 | ... H | R | R | D | K | R | $S_{206}$... | Osatomi and Sumiyoshi, 1990. |
| Dengue virus type 4 | GI: 6978317 | ... R | R | R | E | K | R | $S_{205}$... | Zhao *et al.*, 1986. |
| Japanese encephalitis virus group | | | | | | | | | |
| Japanese encephalitis virus (JaOArS982) | GI: 9626460 | ... S | K | R | S | R | R | $S_{220}$... | Sumiyoshi *et al.*, 1987. |
| Murray Valley encephalitis virus | GI: 9633622 | ... S | K | R | S | R | R | $S_{218}$... | Hurrelbrink *et al.*, 1999. |
| St. Louis encephalitis virus (MSI.7) | GI: 334865 | ... S | R | R | S | R | R | $S_{214}$... | Trent *et al.*, 1987. |
| West Nile virus | GI: 11528013 | ... S | R | R | S | R | R | $S_{216}$... | Castle *et al.*, 1985. |
| Kunjin virus (MRM61C) | GI: 221966 | ... S | R | R | S | R | R | $S_{216}$... | Coia *et al.*, 1988. |
| Rio Bravo virus group | | | | | | | | | |
| Apoi virus (ApMAR) | GI: 7939633 | ... N | T | R | T | R | R | $D_{197}$... | Billoir *et al.*, 2000. |
| Rio Bravo virus (RiMAR) | GI: 7144649 | ... G | H | R | L | K | R | $S_{193}$... | Billoir *et al.*, 2000. |
| Tick-borne encephalitis virus group | | | | | | | | | |
| Langat virus (TP21) | GI: 8453150 | ... G | S | R | S | R | R | $S_{206}$... | Campbell and Pletnev, 2000. |
| Louping ill virus (369/T2) | GI: 9629456 | ... G | S | R | T | R | R | $S_{206}$... | Gritsun *et al.*, 1997. |
| Powassan virus (LB) | GI: 309916 | ... G | S | R | G | K | R | $S_{204}$... | Mandl *et al.*, 1993. |
| Tick-borne encephalitis virus (Neudoerfl) | GI: 9628431 | ... G | S | R | T | R | R | $S_{206}$... | Mandl *et al.*, 1989. |

**Table 1.** (Continued)

| Precursors | GenBank no. | Cleavage site | | | | | | | References |
|---|---|---|---|---|---|---|---|---|---|
| **Flaviviruses prM proteins** | | **... P6** | **P5** | **P4** | **P3** | **P2** | **P1 ↓** | **P1' ...** | |
| Yellow fever virus group | | | | | | | | | |
| Yellow fever virus strain (17D) | GI: 59338 | ... S | R | R | S | R | R | $A_{211}$... | Rice *et al.*, 1985. |
| Unclassified Flavivirus | | | | | | | | | |
| Cell fusing agent | GI: 336190 | ... K | K | R | E | K | R | $S_{220}$... | Cammisa-Parks *et al.*, 1992. |
| **Other viral coat proteins** | | | | | | | | | |
| Avian influenza virus (H5N1) HA | GI: 2833657 | ... R | R | R | K | K | R | $G_{331}$... | Subbarao *et al.*, 1998. |
| Borna disease virus | GI: 15718111 | ... L | V | R | R | R | R | $D_{250}$... | Pleschka *et al.*, 2001. |
| Cytomegalovirus glycoprotein B | GI: 138193 | ... H | N | R | T | K | R | $S_{461}$... | Spaete *et al.*, 1988. |
| Eastern equine encephalomyelitis virus | GI: 2120048 | ... N | A | R | T | R | R | $D_{324}$... | Volchkov *et al.*, 1991. |
| Ebola virus (Zaire) Glycoprotein | GI: 465411 | ... G | R | R | T | R | R | $E_{502}$... | Volchkov *et al.*, 1998. |
| HIV gp160 | GI: 119437 | ... V | Q | R | E | K | R | $A_{507}$... | York-Higgins *et al.*, 1990. |
| Measles virus $F_0$ | GI: 9181897 | ... S | R | R | H | K | R | $F_{116}$... | Parks *et al.*, 2001 |
| Sindbis virus (HRSP) gpE2 | GI: 74511 | ... S | G | R | S | K | R | $S_{329}$... | Strauss *et al.*, 1984. |

The P6-P1' cleavage site sequences for a selected list of proposed subtilisin-like proprotein convertase substrates are shown above. The cleavage sites of these enveloped viruses consist of the highly conserved sequence, -Arg(P4)-Xaa(P3)-Lys/Arg(P2)-Arg(P1)↓-, with basic amino acid motif.

The processing by subtilisin-like proprotein convertase is an important control mechanism for the biological activity of viral surface proteins. Proteolytic activation of envelope glycoproteins is necessary for entry of viruses into the host cell and, hence, for their ability to undergo multiple replication cycles. In some cases, it has also been shown that the cleavability of the envelope glycoproteins is an important determinant for pathogenicity. Proteolytic cleavage is the first step in the activation of these fusion proteins and is followed by a conformational change resulting in the exposure of the fusion domain (Miranda *et al.*, 1996; Stadler *et al.*, 1997; Sabbarao *et al.*, 1998; Volchkov *et al.*, 1998). The conformational change may be triggered by a low pH in the endosomes, as in the case of influenza virus (Sabbarao *et al.*, 1998), or by the interaction with a secondary receptor protein at the cell surface, e.g. in the case of HIV (Miranda *et al.*, 1996).

In the case of influenza virus, the hemagglutinin (HA) is synthesized as a precursor ($HA_0$) that associates noncovalently as homotrimers. The precursor polypeptides are post-translationally cleaved at a conserved arginine residue into two subunits, $HA_1$ and $HA_2$, which linked by a single disulfide bond. This cleavage step is necessary for virus infectivity because it activates the membrane fusion potential of hemagglutinin (Steinhauer, 1999). In 1997, the avian influenza A virus outbreak in Hong Kong, which inflicted a fatal respiratory illness (30% lethality), was traced to an H5N1 strain. Sequence analysis of genes encoding hemagglutinin (HA) from 16 isolates revealed a consistent alteration in the viral genomes to generate a second consensus furin site at the HA1–HA2 junction (-Arg-Lys-Lys-Arg↓-to-Arg-Glu-Arg-Arg-Arg-Lys-Lys-Arg↓-). Exactly how the addition of a tandem consensus site results in enhanced virulence remains to be determined (Suarez *et al.*, 1998; Sabbarao *et al.*, 1998). The mutant avian influenza A virus, in which the amino acid sequence at the HA cleavage site (Pro-Gln-Arg-Glu-Arg-Arg-Arg-Lys-Lys-Arg↓Gly), was converted to sequence of a typical avirulent virus (Pro-Glu- - - - -Arg-Glu-Thr-

Arg↓Gly, where dash indicates a deletion). When tested in mice, this HA mutant was highly attenuated (virus dose lethal to 50% of mice, $MLD_{50}$, $>10^5$ pfu), and none of the infected mice showed signs of disease. The virulence of avian influenza A virus in mice appears to involve HA cleavability (Hatta *et al.*, 2001).

On the other hand, there is currently a debate on subtilisin-like proprotein convertases that are physiologically involved in cleavage and activation of HIV-1 gp160. The production of infectious HIV-1 virions is dependent on the processing of envelope glycoprotein gp160 by the host cell proteinase. The furin and the other subtilisin-like proteases can cleave and activate HIV-1 gp160 to yield gp120 and gp41. This, taken together with furin's expression in $CD4^+$ cell lines, has led to propose that furin is the proteinase activating gp160. However, other proteinases that are subtilisin-like convertase, are also involved physiological in gp160 cleavage (Miranda *et al.*, 1996). Seidah and Chretien (1997) have suggested that furin, SPC6 (PC5/PC6) and SPC7 (LPC/PC7/PC8), are the major gp160-convertase enzymes in T lymphocytes.

The processing of the Ebola virus glycoprotein (GP) by furin was determined by Volchkov *et al* (1998). This was indicated by the observation that cleavage did not occur when GP was expressed in furin-defective LoVo cells, but it was restored in these cells by vector-expressed furin. The Reston subtype, which differs from all other Ebola viruses in its low human pathogenicity, has reduced cleavability due to a mutation at the cleavage site. As a result of these observations, it should now be considered that proteolytic processing of GP might be an important determinant for the pathogenicity of the Ebola virus.