Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis

Q1 Jean Kaoru Millet, Gary R. Whittaker*

1. Introduction

Coronaviruses are a large group of enveloped, single-stranded positive-sense RNA viruses that infect a wide range of avian and mammalian species, including humans. The emergence of deadly human coronaviruses, severe acute respiratory syndrome coronavirus (SARS-CoV), and Middle East respiratory syndrome coronavirus (MERS-CoV) have bolstered research in these viral and often zoonotic pathogens. While coronavirus cell and tissue tropism, host range, and pathogenesis are initially controlled by interactions of the spike envelope glycoprotein and host cell receptor, it is becoming increasingly apparent that proteolytic activation of spike by host cell proteases also plays a critical role. Coronavirus spike proteins are the main determinant of entry as they possess both receptor binding and fusogenic functions. Whereas binding to the host cell receptor is an essential first step in establishing infection, the proteolytic activation step is often critical for the fusion function of spike, as it allows for controlled release of the fusion peptide into target cellular membranes. Coronaviruses have evolved multiple strategies for proteolytic activation of spike, and a large number of host proteases have been shown to proteolytically process the spike protein. These include, but are not limited to, endosomal cathepsins, cell surface transmembrane protease/serine (TMPRSS) proteases, furin, and trypsin. This review focuses on the diversity of strategies coronaviruses have evolved to proteolytically activate their fusion protein during spike protein biosynthesis and the critical entry step of their life cycle, and highlights important findings on how proteolytic activation of coronavirus spike influences tissue and cell tropism, host range and pathogenicity.

* Corresponding author. Tel.: +1 607 253 4019.
E-mail address: gw7@cornell.edu (G.R. Whittaker).

http://dx.doi.org/10.1016/j.virusres.2014.11.021
0168-1702/© 2014 Published by Elsevier B.V.

Please cite this article in press as: Millet, J.K., Whittaker, G.R., Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. Virus Res. (2014), http://dx.doi.org/10.1016/j.virusres.2014.11.021

© 2014 Published by Elsevier B.V.
also enter cells through an endocytic route (Nash and Buchmeier, 1997), whilst the MHV-2 strain enters cells through endocytosis by a clathrin-dependent mechanism (Pu and Zhang, 2008; Qiu et al., 2006).

Importantly, coronaviruses employ a diversity of cues, such as receptor binding, low pH and proteolytic activation, to activate the S protein, allowing a timely release of the fusion peptide into target membranes (Bosch and Rottier, 2008). This enables a spatiotemporally controlled orchestration of viral entry steps.

Along with binding to the host cell receptor, fusion of the viral envelope with host cell membranes is a critical step in establishing successful infection for enveloped viruses such as coronaviruses. The coronavirus S envelope glycoprotein is a class I viral fusion protein (Bosch et al., 2003). The S protein is often activated for fusion by means of proteolytic processing by host cell proteases, an activation process that is typical of class I viral fusion proteins (White et al., 2008). Remarkably, for some coronaviruses, such as SARS-CoV, cleavage of S can occur at two distinct sites (Belouzard et al., 2009). A variety of proteases have been shown to mediate coronavirus activation (Belouzard et al., 2012).

Notably, certain viruses harboring class I viral fusion proteins, like influenza virus and Newcastle disease virus, display characteristic ly expanded or modified cell and tissue tropism, and altered viral pathogenesis following mutation of the cleavage site that results in a change in proteolytic activation (Klenk and Garten, 1994b). This is very well exemplified by the hemagglutinin (HA) protein of highly pathogenic avian influenza (H5N1) virus strains, where transition from a monobasic site, typically cleaved by trypsin-like proteases, to a polybasic site, allows cleavage by ubiquitously expressed furin-like proteases, enabling systemic spread of the virus within an infected host. Thus, small mutational changes in amino acid composition at cleavage sites can have a drastic impact on tissue and cell tropism, host range, and pathogenesis (Klenk and Garten, 1994b; Nagai, 1993).

Here, we put into perspective the wide variety of entry activation mechanisms employed by coronaviruses and present the wide diversity of host cell proteases known to activate coronavirus S proteins. We review what is known about coronavirus proteolytic processing of the S protein and its link with pathogenicity, cell and tissue tropism and host range. We also analyze and compare the amino acid sequence composition of two identified coronavirus S cleavage sites, S1/S2 and S2', in a wide range of coronavirus species encompassing all four coronavirus genera. Finally, we propose using protease sequence recognition motifs on coronavirus S protein as a novel marker to assess pathogenicity and host range, as well as forming the basis for effective therapeutic intervention.

2. Coronavirus S protein

Because they possess an envelope, coronavirus entry into host target cells requires the successful completion of two critical steps. The first is binding to the cell surface by means of attachment to a host cell receptor. The second is fusion of the viral envelope with cellular membranes allowing release of the virus genome into the host cell’s cytoplasm, enabling viral replication to ensue. Both steps are controlled by the S envelope protein (Bosch and Rottier, 2008). As S is a class I viral fusion protein, we will first introduce important basic features of the prototypical class I viral fusion protein, influenza virus HA.

2.1. A prototypical class I fusion protein: Influenza virus hemagglutinin (HA)

There are three classes of virus fusion proteins known, which are classified according to their structural features (Harrison, 2013). While there is a wide degree of variability in the structure and mechanisms involved in the fusion process among these classes, all virus fusion proteins undergo major structural transitions, and ultimately form a final, compact and low-energy trimeric structure, the so-called trimer of hairpins. During the conformational changes, viral and cellular membranes are brought into close proximity. This in turn induces hemifusion, followed by complete fusion of viral and cellular membranes and formation of a pore that can expand, allowing for viral genetic material access into the cell (White et al., 2008).

The very well characterized class I fusion protein influenza virus HA is useful in introducing key structural and mechanistic concepts shared with coronavirus S. Structurally, the salient features of influenza HA are that it assembles in homotrimers orientated perpendicular to the virion membrane surface, has two functional subunits, HA1, which binds to sialic acid receptors and HA2, which contains the fusion machinery, featuring mainly alpha-helical secondary structures (Skehel and Wiley, 2000; White et al., 2008). HA2 contains two heptad repeats, which are structural motifs consisting of chains of seven amino acids that are critical for the fusion function and are another characteristic feature of class I fusion proteins (Chambers et al., 1990). HA is synthesized as an uncleaved precursor named HA0. Fusion activation occurs thanks to endoproteolysis by host cell proteases, such as trypsin-like proteases (Lazarowitz and Choppin, 1975; Lazarowitz et al., 1973a), a cleavage event that processes HA0 into HA1 and HA2 fragments, with both fragments held together by disulfide bonds (White et al., 2008). A fusion peptide, consisting mainly of apolar residues, is found at the N-terminus of HA2 buried at the subunit interface, and is exposed upon cleavage of HA0 and conformational changes of HA1.

A key initial step in the HA fusion process is its priming by proteolytic cleavage that separates HA0 into HA1 and HA2 fragments. At this stage, the HA is in a metastable “spring-loaded” conformation. The receptor-binding subunit HA1 binds to sialic receptors found at the surface of target cells. This triggers uptake and internalization of the virion via the endocytic pathway. During maturation of the endosome, the virion becomes exposed to an increasingly acidic environment (note that this can also occur outside the cell in acidic tissue fluids). This drop in pH is the crucial trigger that provokes further conformational changes of HA allowing for full exposure of the hydrophobic fusion peptide on HA2 (Carr and Kim, 1993): the fusion peptide extends to the tip of the molecule allowing for insertion into the target endosomal membrane, forming an intermediate structure called the prehairpin (White et al., 2008). Further structural rearrangements of several prehairpins occur in which the alpha-helical heptad repeats assemble and bundle up into a compact coiled-coil structure, the six-helix bundle (6HB) with the C-terminal heptad repeats wrapping around the N-terminal heptad repeats (Chen et al., 1999; White et al., 2008). During this dramatic change in structure, viral and target endosomal membranes come into closer proximity. This allows for hemifusion and then full fusion to occur, generating an expanding fusion pore and ultimately allowing release of viral genetic material into the host cell. It is after fusion has occurred, that the HA adopts a structurally stable conformation called the trimer of hairpin.

2.2. The coronavirus spike (S) protein

The coronavirus S protein is a type I transmembrane protein located at the surface of the virion, with a large ectodomain and very short endodomain (Fig. 1) (Masters and Perlman, 2013). As a class I viral fusion protein, it shares many structural and mechanistic features of influenza virus HA (Bosch et al., 2003; Masters and Perlman, 2013). It is the largest of the coronavirus structural proteins with an overall length ranging between ~1200 and 1400 amino acids, and is often heavily glycosylated, with between 21 and 35 N-glycosylation sites.
The S2 domain contains two heptad repeats (HR), which are composed of a chain of seven amino acid patterns, abcdabdc, where a and d are hydrophobic residues, allowing formation of alphahelical structures. HR is a defining structure of class I viral fusion proteins (Fig. 1B). These participate in the formation of coiled-coil structure during membrane fusion. Initially, the precise location of the coronavirus fusion peptide was unclear (Belouzard et al., 2012), however, it is becoming increasingly apparent that the fusion peptide is most likely located adjacent to one of the two cleavage sites, S2' (Burkard et al., 2014; Madu et al., 2009; Masters and Perlman, 2013). This region contains a remarkably conserved motif, I-E-D-L-L-F, which is present in coronavirus S proteins from all four genera (Table 2). Similarly to the influenza HA fusion peptide (Cross et al., 2009), the proposed coronavirus fusion peptide contains mostly hydrophilic residue (I: isoleucine, L: leucine, and F: phenylalanine) with some negatively charged residues (E: glutamic acid, and D: aspartic acid).

The transmembrane domain of the S protein is found in the region C-terminal of the S2 domain (Fig. 1B). A short cytoplasmic tail or endodomain is located at the C-terminal end of S. It contains conserved stretches of cysteine residues, which can be palmitoylated. Palmitoylation of cysteine residues within the endodomain was found to be important for regulating fusogenicity of S (Petit et al., 2007; Shulla and Gallagher, 2009).

Cryo-EM studies on single particles of the SARS-CoV have shed light on the overall architecture of S protein trimers and the conformational changes they undergo during fusion (Benuac et al., 2006; Benuac et al., 2007). While individual structures of key domains of S, such as the receptor binding domain (Li et al., 2005a; Peng et al., 2011; Wu et al., 2009) or the six-helix bundle (Xu et al., 2004), have been extensively studied, it is important to note that, to date, a complete crystal structure determination of the S ectodomain for any coronavirus is still lacking. Such a structure would be very useful to explain accessibility of sites to proteases within S, the precise location of N-linked glycans, and to compare it with other class I viral fusion proteins with known structures, such as influenza virus HA.

3. Coronavirus spike (S) proteolytic activation

3.1. Coronavirus spike (S) can be cleaved at different sites

A peculiar feature of the coronavirus S protein is that it can contain more than one proteolytic cleavage site (Fig. 1B) (Masters and Perlman, 2013). The first identified cleavage site was found to be located at the S1/S2 boundary (Table 1), and another more recently characterized site was found to be within S2 (Table 2), upstream of the putative fusion peptide, and called S2' (Belouzard et al., 2016). S can also be cleaved at other less well-characterized sites. Depending on the coronavirus species and which host cell it infects, cleavage at these different sites can occur at different stages of the virus life cycle, such as during S protein biosynthesis in producer cells and during virus entry into target cells. Cleavage events occurring during biosynthesis or virus entry can be critical factors in modulating pathogenicity, cell and tissue tropism as well as host range. An important distinction from influenza HA is that after cleavage of coronavirus S, the S1 and S2 domains are not held by disulfide bonds, but remain associated non-covalently. Since the two domains are not held covalently, the S1 domain may be shed from the S2 stalk domain of the protein.

Proteases recognize specific amino acid sequences found in their substrates, with the general residue designation: P6–P6' (Fig. 2A). The cleavage site or scissile bond is located between the residue positions P1 and P1'.

Please cite this article in press as: Millet, J.K., Whittaker, G.R., Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. Virus Res. (2014), http://dx.doi.org/10.1016/j.virusres.2014.11.021
### Table 1

Coronavirus spike (S) S1/S2 cleavage sites.

<table>
<thead>
<tr>
<th>Accession no.</th>
<th>S1/S2 sequence</th>
<th>Furin score</th>
<th>Other proteases</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Alphacoronavirus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCoV-Elmo02(I)</td>
<td>AAP72149.1</td>
<td>↓</td>
<td>HTVRRAVRVQTGTTLV83</td>
</tr>
<tr>
<td>FCoV-RM1[1]</td>
<td>ACT10854.1</td>
<td>↓</td>
<td>Q-PRESRSTPN-SV599</td>
</tr>
<tr>
<td>CCoV-171[I]</td>
<td>AAW65515.1</td>
<td>↓</td>
<td>-</td>
</tr>
<tr>
<td>FCoV-1683[II]</td>
<td>AHS58021.1</td>
<td>↓</td>
<td>-</td>
</tr>
<tr>
<td>FCoV-1146[II]</td>
<td>YP_004070194.1</td>
<td>↓</td>
<td>-</td>
</tr>
<tr>
<td>BatCoV-HKU10</td>
<td>YP_00890842.1</td>
<td>↓</td>
<td>-</td>
</tr>
<tr>
<td>PEDV-CT777</td>
<td>AAK38656.1</td>
<td>↓</td>
<td>-</td>
</tr>
<tr>
<td>PEDV-DR13</td>
<td>AFE85969.1</td>
<td>↓</td>
<td>-</td>
</tr>
<tr>
<td>TGEV-Miller</td>
<td>ABG89306.1</td>
<td>↓</td>
<td>-</td>
</tr>
<tr>
<td>TGEV-Purdue</td>
<td>ABG89335.1</td>
<td>↓</td>
<td>-</td>
</tr>
<tr>
<td>HCoV-NL63</td>
<td>AASS8177.1</td>
<td>↓</td>
<td>-</td>
</tr>
<tr>
<td>HCoV-229E</td>
<td>BAL46537.1</td>
<td>↓</td>
<td>-</td>
</tr>
<tr>
<td><strong>Betacoronavirus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lineage A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCoV-Quebec</td>
<td>P25193.2</td>
<td>↓</td>
<td>STKRRRSRSTTGYYF776</td>
</tr>
<tr>
<td>HCoV-OC43</td>
<td>AAA03055.1</td>
<td>↓</td>
<td>KSRRSRGAITGYYRF766</td>
</tr>
<tr>
<td>HCoV-HKU1</td>
<td>AAT38580.1</td>
<td>↓</td>
<td>SSSRRRSRISASRYF68</td>
</tr>
<tr>
<td>MVH-AS9</td>
<td>AAAAA6455.1</td>
<td>↓</td>
<td>SKRRRARSSVTSTGYYL720</td>
</tr>
<tr>
<td>MVH-JHM</td>
<td>YP_209233.1</td>
<td>↓</td>
<td>SKRRRARSSVTSTGYYL77</td>
</tr>
<tr>
<td>Lineage B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV</td>
<td>NP_828851.1</td>
<td>↓</td>
<td>HTYSLR1SSQ877</td>
</tr>
<tr>
<td>Lineage C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MERS-CoV</td>
<td>AFS88936.1</td>
<td>↓</td>
<td>TTPRSPRSVPSGEMRL799</td>
</tr>
<tr>
<td>BatCoV-HKU4</td>
<td>YP_00139953.1</td>
<td>↓</td>
<td>VSTTFFYSASQ-----755</td>
</tr>
<tr>
<td>BatCoV-HKU5</td>
<td>YP_00139962.1</td>
<td>↓</td>
<td>TTSSRVRATSGADY753</td>
</tr>
<tr>
<td><strong>Gammacoronavirus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBV-Beaudette</td>
<td>NP_040831.1</td>
<td>↓</td>
<td>TTRRSSTVANSA5</td>
</tr>
<tr>
<td>IBV-M41</td>
<td>AAW37787.1</td>
<td>↓</td>
<td>TTRRSSTVANSA5</td>
</tr>
<tr>
<td>IBV-Ca99</td>
<td>AASS0080.1</td>
<td>↓</td>
<td>THERSRENNVTSTT54</td>
</tr>
<tr>
<td>TurkeyCoV-MG10</td>
<td>ABW81427.1</td>
<td>↓</td>
<td>SRKRRSTNAVYTGCT53</td>
</tr>
<tr>
<td>BeCoV-SW1</td>
<td>YP_001876437.1</td>
<td>↓</td>
<td>HVGRDVSVNTGVNCH813</td>
</tr>
<tr>
<td>BdCoV-HKU22</td>
<td>AHB63508.1</td>
<td>↓</td>
<td>QAALS-EEEQVNCN818</td>
</tr>
<tr>
<td><strong>Delta-coronavirus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BuCoV-HKU11</td>
<td>ACJ12035.1</td>
<td>↓</td>
<td>SCKTERTIAASTFY27</td>
</tr>
<tr>
<td>ThCoV-HKU12</td>
<td>ACJ12053.1</td>
<td>↓</td>
<td>TILN-FKRATPTFY554</td>
</tr>
<tr>
<td>MuCoV-HKU13</td>
<td>ACJ12062.1</td>
<td>↓</td>
<td>TLYC-FQRPMTFY119</td>
</tr>
</tbody>
</table>

2. Li et al. (2006).

#### 3.2. Cellular proteases involved in coronavirus activation

**3.2.1. Trypsin**

Trypsin (Fig. 2B) is a prototype serine endopeptidase and has been studied extensively in the context of virus glycoprotein cleavage-activation. Trypsin cleaves at neutral pH, with an optimum of approximately 8.0. Biochemically, trypsin strongly prefers to cleave at arginine (R) or lysine (K) residues, although almost all viral substrates get cleaved at arginines. Other flanking residues are in general non-discriminatory, however basic or hydrophobic residues at P2 decreases the catalytic activity, as does a proline (P) residue at P3 (Hallon et al., 2004). Overall the requirement for cleavage is for a single basic residue at P1. As trypsin is not very selective in terms of substrate recognition, there are potentially many sites within coronavirus S proteins that may be cleaved by it. However, within the native trimeric conformation of S, only a few of these sites, if any, are accessible to trypsin cleavage. Trypsin is expressed in the respiratory tract, where its activity is inhibited by alpha-1 antitrypsin. Trypsin is principally a digestive enzyme, with active trypsin found in the small intestine; as such trypsin is likely to be able to directly cleave the S proteins of many enteric coronaviruses. Probably the best example of this is porcine epidemic diarrhea virus (PEDV), where most strains are highly trypsin-dependent. In the case of PEDV produced in Vero cells, trypsin cleavage only occurs after receptor binding, and not on free particles (Park et al., 2011a). For those coronaviruses with respiratory tissue tropism, cell-culture based studies using trypsin likely act as a surrogate for more biologically relevant proteases such as members of the TMPRSS family, which have similar substrate specificities. A good example of this is SARS-CoV produced in VeroE6 cells where trypsin can override the need for cathepsin-mediated cleavage, and shifts the virus to a low-pH independent route of entry, possibly at the...
Table 2
Coronavirus spike (S) 2′ cleavage sites.

<table>
<thead>
<tr>
<th>Accession no.</th>
<th>S2′ sequence</th>
<th>Furin score</th>
<th>Other proteases</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCoV-Elmo/02(I)</td>
<td>AAP72149.1</td>
<td>↓&lt;sub&gt;65&lt;/sub&gt;LPAPCGRGRRSAIEDLF&lt;sub&gt;1000&lt;/sub&gt;</td>
<td>&lt;0</td>
</tr>
<tr>
<td>FCoV-RM1</td>
<td>ACT10854.1</td>
<td>↓&lt;sub&gt;65&lt;/sub&gt;LPTTFGRGRRSAIEDLF&lt;sub&gt;986&lt;/sub&gt;</td>
<td>&lt;0</td>
</tr>
<tr>
<td>CCoV-1-71(B)</td>
<td>AAV65515.1</td>
<td>↓&lt;sub&gt;65&lt;/sub&gt;HNSRKRYSRSAIEDLF&lt;sub&gt;970&lt;/sub&gt;</td>
<td>+5.3</td>
</tr>
<tr>
<td>FCoV-1683(B)</td>
<td>AHF58021.1</td>
<td>↓&lt;sub&gt;65&lt;/sub&gt;HNSRKRYSRSAIEDLF&lt;sub&gt;957&lt;/sub&gt;</td>
<td>+5.3</td>
</tr>
<tr>
<td>FCoV-1146(B)</td>
<td>YP0004709194.1</td>
<td>↓&lt;sub&gt;65&lt;/sub&gt;HNSRKRYSRSAIEDLF&lt;sub&gt;950&lt;/sub&gt;</td>
<td>&lt;0</td>
</tr>
<tr>
<td>BatCoV-HKU10</td>
<td>YP00089642.1</td>
<td>↓&lt;sub&gt;65&lt;/sub&gt;KDGK5V9VEDLF&lt;sub&gt;73&lt;/sub&gt;</td>
<td>&lt;0</td>
</tr>
<tr>
<td>PEDV-CV777</td>
<td>AAF58356.1</td>
<td>↓&lt;sub&gt;85&lt;/sub&gt;SRVRQGVRQGVRQG&lt;sub&gt;1140&lt;/sub&gt;</td>
<td>+7.4</td>
</tr>
<tr>
<td>PEDV-DR13</td>
<td>ADE38596.1</td>
<td>↓&lt;sub&gt;85&lt;/sub&gt;SRVRQGVRQGVRQG&lt;sub&gt;1140&lt;/sub&gt;</td>
<td>&lt;0</td>
</tr>
<tr>
<td>TGEV-Miller</td>
<td>ABG89306.1</td>
<td>↓&lt;sub&gt;85&lt;/sub&gt;DNSRKRYSRSAIEDLF&lt;sub&gt;963&lt;/sub&gt;</td>
<td>+5.3</td>
</tr>
<tr>
<td>TGEV-Purdue</td>
<td>ABG89335.1</td>
<td>↓&lt;sub&gt;85&lt;/sub&gt;HNSRKRYSRSAIEDLF&lt;sub&gt;966&lt;/sub&gt;</td>
<td>+5.3</td>
</tr>
<tr>
<td>HCoV-NL63</td>
<td>AAS58177.1</td>
<td>↓&lt;sub&gt;85&lt;/sub&gt;RSSRAGRSRSAIEDLF&lt;sub&gt;978&lt;/sub&gt;</td>
<td>+8.6</td>
</tr>
<tr>
<td>HCoV-229E</td>
<td>BAI5637.1</td>
<td>↓&lt;sub&gt;85&lt;/sub&gt;SRSRAGRSRSAIEDLF&lt;sub&gt;978&lt;/sub&gt;</td>
<td>+8.6</td>
</tr>
</tbody>
</table>

**Alphacoronavirus**

**Lineage A**
- CoV-Quebec: P25193.2
  - ↓<sub>90</sub>DCRVSSSSRSASG<sub>101</sub> | +0.6 |
- HCoV-OC43: AAA03055.1
  - ↓<sub>85</sub>KCSPSSASG<sub>911</sub> | <0 |
- HCoV-HKU1: AAT98580.1
  - ↓<sub>85</sub>KHCSSSSASG<sub>912</sub> | <0 |
- MHV-A59: AAI4744.1
  - ↓<sub>85</sub>GPSAIGRSASG<sub>911</sub> | <0 |
- MHV-BJ: YP_2020233.1
  - ↓<sub>85</sub>GPSAIGRSASG<sub>911</sub> | <0 |

**Lineage B**
- SARS-CoV: NP_828851.1
  - ↓<sub>70</sub>DPALKPRTKS<sub>805</sub> | <0 |
- Elastase<sup>2</sup>, trypsin<sup>3</sup>

**Betacoronavirus**

**Lineage A**
- MERS-CoV: NP_828851.1
  - ↓<sub>70</sub>DPALKPRTKS<sub>805</sub> | <0 |
- BatCoV-HKU4: YP_001039955.1
  - ↓<sub>85</sub>GGSSSRSASG<sub>94</sub> | +8.6 |
- BatCoV-HKU5: YP_001039962.1
  - ↓<sub>85</sub>TTGERKYSTTEDLF<sub>92</sub> | <0 |

**Gammacoronavirus**

- IBV-Beaudette: NP_004831.1
  - ↓<sub>65</sub>NPSRKRYSR<sub>19</sub> | +12.7 |
- IBV-M4: AAA17376.1
  - ↓<sub>65</sub>TPSPRKRYSR<sub>19</sub> | <0 |
- IBV-Ca99: AAB50080.1
  - ↓<sub>65</sub>TPSPRKRYSR<sub>19</sub> | <0 |
- TurkeyCoV-MG10: ABBW1427.1
  - ↓<sub>65</sub>TLAQGCRSTTEDLF<sub>112</sub> | <0 |
- BeCoV-SW1: YP_001878437.1
  - ↓<sub>65</sub>GDPRDARSTTEDLF<sub>981</sub> | <0 |
- BdCoV-HKU2: AHE63508.1
  - ↓<sub>65</sub>SGDPRDARSTTEDLF<sub>1004</sub> | +1.3 |

**Deltacoronavirus**

- BuCoV-HKU11: ACJ12035.1
  - ↓<sub>65</sub>ITSKSGGRSR<sub>860</sub> | <0 |
- ThCoV-HKU12: ACJ12035.1
  - ↓<sub>65</sub>ITSKSGGRSR<sub>860</sub> | <0 |
- MuCoV-HKU13: ACJ12062.1
  - ↓<sub>65</sub>ITSKSGGRSR<sub>860</sub> | <0 |

1. Wicht et al. (2014b).
2. Belouard et al. (2010).
4. membrane plasma. As with PEDV, prior binding to the viral receptor is a pre-requisite for SARS-CoV fusion activation (Matsuyama et al., 2005). In other cases, such as with feline coronavirus (FCoV), trypsin can activate the virus without the need for prior engagement of the viral receptor (Costello et al., 2013).

3.2.2. Furin and the proprotein convertase (PC) family

A large number of coronaviruses, from different genera, possess at their S1/S2 boundary site a furin cleavage site (Table 1). Furin (Fig. 2B) and furin-like proteases cleave at paired basic residues, and belong to the group of proteases called proprotein convertases. The PC family is comprised of nine serine proteases that can cleave a vast number of cellular and microbial protein substrates. Furin belongs to the subset of PCs that includes proconvertase 1 (PC1), PC2, PC4, PC5, paired basic amino acid cleaving enzyme 4 (PACE4) and PC7, which generally cleave at single or paired basic residues within the following motif X-K/-R/-X<sub>2,4,6</sub>-K/-R/-X<sub>2,4,6</sub>-K/-X<sub>2,4,6</sub> (X: any amino acid) (Seidah and Prat, 2012). The other two PCs, subtilisin/kexin-3 isoform-1 (SKI-1) and proprotein convertase subtilisin/kexin type 9 (PCSK9) cleave at non-basic residues sites. SKI-1 recognizes R-X<sub>2</sub>-L-V/I/-Y motifs (L: leucine, V: valine, I: isoleucine) and PCSK9 only autoprotoproteolytically cleaves itself at the V<sub>F</sub>-A-Q motif (F: phenylalanine, A: alanine, Q: glutamine). PCs are crucial regulatory proteases as they can activate and sometimes inactivate a very wide variety of substrates, such as enzymes and other proteases, cell adhesion factors, growth factors and hormones. The critical role of PCs is further demonstrated by knockout mouse experimental models, where individual knockout of furin, PC5, PACE4 or SKI-1 are lethal at different stages of development (Seidah and Prat, 2012).

PCs are known to be subverted by pathogenic microorganisms as they are recruited by viruses to process viral proteins, such as HIV gp160, respiratory syncytial virus (RSV) F protein, human papilloma virus minor capsid protein L2, dengue virus prM, and Chikungunya virus E3E2 (McCune et al., 1988; Ozden et al., 2008; Richards et al., 2005; Rodenkvist-Syburt et al., 2010; Zimmerman et al., 2003). They are also implicated in the cleavage of bacterial toxins, e.g. Pseudomonas exotoxin and Shiga toxin (Molloy and Thomas, 2001).

Because furin can be viewed as a representative PC (Thomas, 2002), and because it is the most frequently characterized PC to be involved in processing and activating coronavirus S proteins, we will focus on this protease. However, it is important to recognize that there is a degree of redundancy between substrate recognition motifs by the different PCs, especially between furin, PC5 and PACE4. In line with this, a study has shown that overexpression of PACE4 and PC5 could enhance proteolytic cleavage of SARS-CoV S, along with furin (Bergeron et al., 2005).

Like all PCs, furin is a serine protease, containing a critical serine residue in its catalytic triad. Also, like all PCs, it is related to the bacterial subtilisin family of protease, in which the prototypical yeast protease kexin is found. PCs require Ca<sup>2+</sup> ions to be active. While most PCs, including furin require neutral pH for activity, a study has shown that self-activation of furin via protonation of a regulatory...
Fig. 2. Host cell proteases involved in activating the coronavirus spike (S) protein. (A) Schematic of a protease cleavage site and substrate binding pocket. The sites within the protease that accommodate substrate residues are designated with the letter S. The residues of the substrate protein involved in recognition and proteolytic processing are denoted with the letter P. The scissile bond is cleaved by the protease and the residues involved in this bond are denoted P1–P1'. (B) Structures of three common host cell proteases known to activate coronavirus S: crystal structures of trypsin (PDB: 2PTN), furin (PDB: 1PB), and the pro-form of cathepsin L (PDB: 1CJL). (C) Diagram of a coronavirus life cycle and the various host cell proteases known to cleave and activate some coronavirus S proteins. Note that for certain coronaviruses, fusion can occur directly at the plasma membrane.

Histidine residue (His-69) requires a slightly acidic environment (Williamson et al., 2013). The minimal substrate recognition motif for furin is R-X-X-R, with the R-X-R/K-R motif being highly favorable. The fairly strict requirement for basic arginines at P1 and P4 positions as well as lysine at P2 is due to furin's binding pocket containing complementary charged residues (Henrich et al., 2003). Also, flanking serine (S) residues in the vicinity of the cleavage site appear to be highly favored by furin.

Several computational tools, such as ProP 1.0 and PiTou 2.0, have been made available for users to query amino acid sequences and predict whether or not they can be cleaved by furin (Duckert et al., 2004; Tian et al., 2012). The PiTou 2.0 tool offers the advantage of analyzing the 20 residue recognition motif used by furin (Tian, 2009; Tian et al., 2011), and takes into account the binding strength and solvent accessibility of substrate residues, allowing for high sensitivity and specificity predictions (Tian et al., 2012).

Furin is widely expressed and is often considered as a ubiquitous protease. However it should be noted that in most cases, the levels of furin expression are low (Shapiro et al., 1997). Furin is produced in the ER, and traffics along the secretory pathway, where it can process coronavirus S proteins in infected cells. Furin is a membrane bound protease that can be shed in the extracellular space, is constitutively secreted, and is mainly found in the trans-golgi network (TGN) where it is activated and can traffic to the plasma membrane. Furin can also recycle back from the plasma membrane to the TGN through an endocytic pathway (Seidah and Prat, 2012). It is noteworthy that while most coronaviruses that are cleaved by furin or furin-like proteases are cleaved at only one site, the S1/S2 site, there are examples of coronaviruses, such as IBV and MERS-CoV that can be additionally cleaved at the putative fusion peptide-proximal site S2' (Table 2), as described in Section 2.4 Proteolytic activation mechanisms of coronaviruses.

3.2.3. Cathepsins

The cathepsins comprise a group of cysteine, serine, and aspartyl proteases with both endo- and exo-peptidase activity that are typically found in endosomes and lysosomes and function as degradative enzymes, as well as in antigen processing.

Cathepsins may also be secreted however, especially in cancer cells (Mohamed and Sloane, 2006). As a general rule, the substrate specificity of cathepsins is quite broad, and analysis of the residues recognized by these enzymes seems to differ depending on the analyses used. Biochemical analysis of peptide substrates indicates that arginine (R) residues are preferred at the P1 position, often with
an aromatic residue at P2 (Choe et al., 2006; Polgár, 1989; Rawlings and Barrett, 2004). However, databases of natural cathepsin sub-
strates (e.g. MEROPS) show a preference for non-polar residues
such as glycine (G), alanine (A), serine (S), threonine (T), and glu-
tamine (Q) at the P1 position, along with a bulky hydrophobic residue at P2. Given their predominant endosomal/lysosomal loca-
tion there is a distinct requirement of low pH for enzymatic activity.

Cathepsin L (Fig. 2B) is most commonly associated with activation
of viral glycoproteins, and has a pH optimum of 3.0 to 6.5. Cathepsin
L is known to process SARS-CoV S, as well as a range of other coro-
naviruses, including MERS-CoV, HCoV-229E, and MHV-2 (Kawase
et al., 2009; Qi et al., 2006; Shirato et al., 2013; Simmons et al.,
2005). Cathepsin B can also be involved in coronavirus entry, and
has some distinct properties compared to cathepsin L and other
cathepsins; it has a generally higher pH optimum and, while pre-
ferring an aromatic P2 residue, can process di-basic substrates, in
contrast to the other cathepsins (Choe et al., 2006; Mort, 2004;
Polgár, 1989). Coronavirus using cathepsin B for entry include
type II feline coronavirus (FCoV) and MHV-2 (Qi et al., 2006; Regan
et al., 2008).

3.2.4. Other proteases: Transmembrane protease/serine (TMPRSS)
proteases, elastase, plasmin

Members of the TMPRSS family are membrane-bound trypsin-
like serine proteases that are found on the cell surface or in the
secretory pathway of cells (Antais et al., 2011; Simmons et al.,
2013). TMPRSS proteases are part of the larger family of proteases,
including the type II transmembrane serine protease or TTSP (Bugge et al.,
2009). TTSPs are subdivided into four groups: the HAT/DESC, Hep-
sin/TMPRSS, Matripsase, and Corin subfamilies. TMPRSS proteases
are type II transmembrane proteases. They are widely expressed in
the respiratory tract and are involved in the activation of many
respiratory viruses. Both MERS–CoV S and SARS-CoV S can be ac-
tivated by TMPRSS family members, as well as HCoV-229E S (Bertram
et al., 2013; Gierer et al., 2013; Glowacka et al., 2011; Shirato et al.,
2013). Intriguingly, TMPRSS2 was shown to be important for the
late stages of PEDV life cycle (Shirato et al., 2011). While the sub-
strate specificity of TMPRSS proteases is not well defined, it is
generally thought to be similar to that of trypsin.

Recently, the TTSPs mosaic serine protease large form (MSPL),
and differentially expressed in squamous cell carcinoma 1 (DESC1)
were shown to activate MERS-CoV and SARS-CoV S proteins for
cell–cell and virus–cell fusion (Zmora et al., 2014). MSPL and DESC1
were also shown to activate the HA proteins of influenza virus. The
authors demonstrated by quantitative RT–PCR that while MSPL was readily expressed in human lung tissue, DESC1 was only expressed
at low levels. Remarkably, TMPRSS13, a splice variant form of MSPL,
was shown to be sensitive to inhibition by the naturally occurring
Kunitz-type serine protease inhibitor, hepatocyte growth factor
activator inhibitor type 1 (HAI-1) (Hashimoto et al., 2010). HAI-
1 was shown earlier to be involved in the regulation and inhibition
of TTSPs such as hepsin, matripsase and prostatin (Fan et al., 2005;
Kirchofer et al., 2005; Lin et al., 1999). A related protease inhibitor,
HAI-2 was shown to inhibit influenza virus infection in vitro and its
administration protected animals in a mouse model for influenza
infection (Hamilton et al., 2014). It would be of interest to test such
naturally occurring inhibitors for their effects on coronavirus entry
and infection.

Elastase has been shown to activate SARS-CoV S. While the
biological relevance of this has focused on elastase produced by
neutrophils during the inflammatory response, most experiment-
mental studies have used pancreatic elastase. Like trypsin, pancreatic
elastase can shift SARS-CoV entry to a low pH-independent route
(Belouzard et al., 2010; Matsuyama et al., 2005), although the amino
acid preferences for elastase are very different to trypsin, with
alanine (A), serine (S), glycine (G) and valine (V) residues being preferred at the P1 position (Bieth, 2004).

Plasmin is a key enzyme in blood clot lysis and its major natu-
ral substrates are fibrinogen and fibrin (Castellino, 2004; Castellino
and Ploplis, 2003). It is produced from its precursor plasminogen.
Plasmin is also involved in a variety of other cellular processes in
various tissues. Like trypsin, plasmin cleaves at K- and R-bonds,
but is much less efficient and cleaves only a subset of such bonds.
In part, this is due to differences in the P2 residues, where bulky
hydrophobic residues are preferred (Bakkes et al., 2000; Hervio
et al., 2000). Thus plasmin is a much more selective enzyme than
trypsin. Plasmin is known to activate certain influenza virus HAs
(Lazarowitz et al., 1973b; Sun et al., 2010; Tse et al., 2013). Although
plasmin has been shown to activate SARS-CoV S (Kam et al., 2009),
there remains little direct biological evidence for a role of plasmin
in activating SARS-CoV or other coronaviruses.

Another possible source of proteases that may impact corona-
virus pathogenesis is proteases produced by co-infecting bacteria.
Many bacterial species secrete proteases as virulence factors, or
elicit secretion of host proteases, and these proteases are likely to
have the capacity to activate virus envelope glycoproteins under
certain co-infection conditions. Such a situation has been explored
for influenza virus (Böttcher-Friebertshäuser et al., 2013), but has
not generally been evaluated in coronaviruses, except for infection
of mice with SARS-CoV and Pasteurella pneumotropica, a low-
pathogenicity bacterium that elicits elastase production in the
lungs (Ami et al., 2008).

3.3. Spatio–temporal regulation of coronavirus fusion

One important component of S protein activation is that pro-

tease action is integrated into the life cycle of the virus, with
cleavage able to occur during biosynthesis of the S protein during
virus assembly, in the extracellular space on released particles,
or on the cell surface and in endosomal compartments during
virus entry (Fig. 2C). As a general rule, cleavage during biosyn-
thesis is likely to be carried out by furin and related PCs that have
steady-state localization to the ER/Golgi and trans-Golgi compart-
ments. Cleavage by trypsin and TMPRSS family members would
be expected to occur in the extracellular space and cell surface,
respectively. Cleavage during virus entry is expected to be mediated
by cathepsins. However, it is important to note that intracellu-
lar proteases cycle between membrane compartments, and many
proteases can shed catalytically active ectodomains from the cell
surface. For instance, TMPRSS proteases can be active in the Golgi,
furin can be present on the cell surface or in endosomes, and some
cathepsins can be active in the secretory pathway or outside the
cell. Such spatio-temporal protease cleavage events can be modu-
lated extensively in different cell types and by the physiological
status of a given cell type.

3.4. Proteolytic activation mechanisms of coronaviruses

3.4.1. Severe acute respiratory syndrome coronavirus (SARS-CoV)

The emergence of SARS-CoV in the human population reinforced
research in coronavirus biology. In particular, the entry pathways
and activation of SARS-CoV, a betacoronavirus, have become the
subject of intense study. The proteolytic activation mechanisms
SARS-CoV S protein are perhaps amongst the best characterized so
far and provide a good model to investigate coronavirus S proteoly-
tic processing because many details of the proteases involved and
cleavage sites processed are known. SARS-CoV also represents an
example of how virus envelope glycoproteins can be activated by
a diverse array of mechanisms.

Initially, a study employing retroviral pseudovirions produced in
HEK-293T cells and harboring SARS-CoV S has found that

Please cite this article in press as: Millet, J.K., Whittaker, G.R., Host cell proteases: Critical determinants of coronavirus tropism and
SARS-CoV entry was sensitive to lysosomotrophic agents, indicating a role for endosomal low pH (Simmons et al., 2004). Using similar HEK-293T produced SARS-CoV S pseudovirions, it was later shown that the dependency of virus entry on low-pH was because SARS-CoV S can be activated by endosomal cathepsin L, which requires low pH for its enzymatic activity (Simmons et al., 2005).

Surprisingly, SARS-CoV seems to have evolved a multi-pronged approach regarding the activation of its S as many other proteases were shown to be able to convert S-mediated entry and fusion, most of which act on the virus outside of the cell. Cathepsin L was shown to cleave SARS-CoV S at residue T678 in the S1/S2 boundary (Table 1) (Bosch et al., 2008). Trypsin was shown to activate S-induced syncytia formation, without the need to acidify the extracellular milieu (Simmons et al., 2004), indicating that the low pH requirement observed for SARS-CoV entry is not due to S requiring low pH for fusion competency. Trypsin was found to cleave SARS-CoV S at R667 in the S1/S2 site (Table 1) (Li et al., 2006).

Furthermore, using native SARS-CoV infection in VeroE6 cells, it was shown that along with trypsin, thermolysin and elastase were also able to activate SARS-CoV fusion and entry (Matsuyama et al., 2005). The authors showed that cleavage activation by trypsin and thermolysin on virions bound on the cell surface allowed a much more efficient entry (>100-fold enhancement of infectivity) than endosomal cathepsin-mediated entry. Notably, prior treatment of virions before cell attachment decreased infectivity. This result suggests that fusion-active cleaved SARS-CoV S proteins are unstable, and the conformational changes induced by the cleavage event are transient and irreversible.

The location of the cleavage sites within SARS-CoV S has been investigated. Using wild type and mutated SARS-CoV S pseudovirions produced in HEK-293T cells, it was found that S could be cleaved by trypsin at two distinct sites, one located at the boundary of S1 and S2, the “classical” S1/S2 site (R667 P1 residue), and another newly identified site, S2’ (R797 P1 residue), found upstream of the putative fusion peptide and within the S2 domain (Tables 1 and 2) allowing for activation of fusion and entry (Belouzard et al., 2009).

Importantly, trypsin cleavage of SARS-CoV S is thought to be sequential, with the S1/S2 cleavage occurring first and enhancing subsequent cleavage at S2’. It is the second cleavage event, at S2’, that is believed to be crucial for fusion activation of S. The S1/S2 cleavage appears dispensable for syncytia formation and virus–cell fusion.

Elastase-mediated cleavage was shown to occur at the S2’ site of SARS-CoV S on pseudovirions produced in HEK-293T cells, with the P1 residue found to be a threonine, T795 (Table 2) (Belouzard et al., 2010). Cleavage at this site shifts the starting position of the putative fusion peptide, which modulates but does not abrogate fusogenicity of S. A study has shown that plasmin, a known protease activator of influenza HA, could, along with trypsin, cleave SARS-CoV S expressed in BHK–21 cells (Kam et al., 2009).

The respiratory-tract resident and cell surface-expressed TMPRSS2 and TMPRSS11a proteases were shown to cleave and mediate activation of SARS-CoV entry in studies using HEK-293T-produced SARS-CoV S pseudoviruses (Glowacka et al., 2011; Kam et al., 2009). This is a feature shared with influenza HA, which is known to be activated by TMPRSS proteases (Böttcher et al., 2006).

Remarkably, it was found that TMPRSS2 associates with the SARS-CoV receptor, ACE2, forming a receptor-protease complex allowing for efficient entry of the virus, directly at the cell surface (Shulla et al., 2011). Another TMPRSS protease, TMPRSS11d also known as human airway trypsin-like protease (HAT), was also shown to proteolytically activate SARS-CoV S in a HEK-293T expression system (Bertram et al., 2011). Using mass spectrometry, the authors demonstrated that TMPRSS11d cleaved S at R667, in the S1/S2 cleavage site (Table 1). While TMPRSS2 cleavage appears to occur at different sites within S, TMPRSS11d cleavage occurs mostly at the S1/S2 site (R667), which may explain why TMPRSS11d activation appears sufficient for cell–cell fusion but not virus–cell fusion. After TMPRSS2 cleavage, shed fragments of S, including a large 150 kDa fragment, are detected in the supernatant of cells and were shown to act as decoys to anti-S neutralizing antibodies (Glowacka et al., 2011).

Factor Xa is yet another protease that was shown to activate SARS-CoV S for entry into host cells (Du et al., 2007). The authors showed that the S protein of SARS-CoV retroviral pseudovirions produced in HEK-293T cells could be cleaved by factor Xa at the S1/S2 boundary. This cleavage event occurred when the pseudovirions were incubated with susceptible cells that were shown to express the protease.

Overall, the study of SARS-CoV S was instrumental in defining cleavage sites and putting forward the notion that the coronavirus S protein can be cleaved by a wide variety of proteases, suggestive of a relatively high degree of plasticity in cleavage mechanisms employed.

3.4.2. Murine hepatitis virus (MHV)

The betacoronavirus MHV has been extensively studied and represents a good model for coronavirus biology and in particular for studying coronavirus entry mechanisms. In groundbreaking work on MHV strain A59 (MHV-A59), the S protein found on virions (then known as E2), was shown to be cleaved during its biosynthesis when produced in four infected murine cell lines: 17 Cl 1, Sac‘, DBT and L2 (Frana et al., 1985). Pulse-chase experiments demonstrated that the cleavage event occurred before virion release from productively infected cells. The authors showed that the degree of cleavage varied depending on which cell type the S protein was expressed in. A study from the same group showed that the 180 kDa precursor S protein of MHV-A59 virions produced in the 17 Cl 1 mouse cell line was cleaved approximately in two halves, with equal amounts of uncleaved and cleaved S and yielding two ~90 kDa cleavage products (Sturman et al., 1985). The proteolytic processing of S was found to be potentiated by trypsin treatment of virions produced in murine 17 Cl 1 cells. The site where trypsin cleaved MHV-A59 S was mapped to the S1/S2 site (Table 1) (Luytjes et al., 1987). It was later shown that MHV-A59 S is cleaved by furin during biosynthesis and trafficking of S through the trans–Golgi network in murine LR-7 cells (de Haan et al., 2004). MHV-A59 S contains a minimal furin cleavage motif (R–R–A–H–R–S), containing at least P1 and P4 arginine residues (R–X–X–R) at the S1/S2 boundary, a feature that is shared by many alphacoronaviruses and many other coronaviruses from the beta- and gamma-genera (Table 1). Mutating the S1/S2 site was shown to greatly affect the degree of cleavage of the S protein (de Haan et al., 2004). Another important concept that comes out of MHV studies is that the degree of cleavage at S1/S2 found when S is expressed and matures in different cell lines reflects the varying abundance of protease expression.

The MHV-2 strain was found to have lost this cleavage site and its S is uncleaved on virions exiting murine DBT producer cells (Yamada et al., 1997). When expressed in cells, MHV-2 fails to induce syncytia at neutral pH. MHV-2 S-induced syncytia can form if exogenous protease treatment is applied. It appears that MHV-2 still requires proteolytic processing for fusion competency. Indeed, MHV-2 S-mediated entry of virions produced in 17 Cl 1 mouse cells is sensitive to inhibitors of the endosomal compartment resident proteases cathepsin B or L in murine fibroblast L2 target cells (Qiu et al., 2006). This suggests that for MHV-2, the timing and nature of proteases involved has been switched from furin-like proteolytic processing during virus maturation in producer cells to endosomal cathepsins during entry in target cells.

Recent work reported that the S protein of MHV-A59, in addition to being cleaved at S1/S2 during protein biosynthesis in mouse LR7 producer cells, is cleaved a second time at a site near the putative
fusion peptide, named S2*, during the viral entry process (Wicht et al., 2014a). This site is analogous to the S2* site in SARS-CoV S (Belouzard et al., 2009). The authors devised a clever experimental setup using conditional biontintylation assay that allows specific labeling and detection of MHV S proteins that have undergone the fusion process. This allows the analysis of any cleavage activation that may have occurred before fusion takes place. A new ~80 kDa band was detected by Western blot analysis, which the authors conclude to be the fusion subunit. A comprehensive study, using siRNA screenings and a replication-independent fusion assay, has shown that MHV-A59 entry necessitates clathrin-mediated endocytosis and that the virions fuse with lysosomal membranes (Burbard et al., 2014). The study also conclusively demonstrates that MHV-A59 requires cleavage of S, at the S2* site. Inhibition of MHV-A59 entry could be obtained by using a pan-lysosomal protease inhibitor. The identity of the specific protease involved awaits elucidation.

3.4.3. Bovine coronavirus (BCoV)

A recent study on the betacoronavirus BCoV has uncovered important features of coronavirus population adaptation during host, tissue, and cell tropism shifts (Borucki et al., 2013). Intriguingly, the authors found a critical role for a genotype containing a S variation at the S2* site that rapidly expands and dominates the population during the course of passaging in cell culture (Table 3). Using a deep-sequencing approach the authors looked at the effect of passaging field strains of BCoV obtained in nasal washes on the genotype composition of the population. The viruses were passaged in human or bovine macrophage cell lines (THP-1 and BOMAC) as well as in a human rectal cell line (HRT-18). Surprisingly, they found that a variant genotype, a minority present in unp passaged samples, and which harbored a mutated S containing a multibasic insert composed of three arginines at S2*, was independently selected during passaging of three different nasal wash samples in either BOMAC, THP-1 or HRT-18 cell lines (for example, the B2.27.Bo.P1 BOMAC passaged virus, Table 3). Surprisingly, the study showed that just a single passaged in BOMAC cells was sufficient to expand the minority variant to become dominant in the viral population. It is notable that the insertion of the three basic arginines (R-R-R) is very similar to what is observed in the HA polybasic sites of highly pathogenic avian influenza virus subtypes H5 and H7 (Kawaoka and Webster, 1988; Klent and Garden, 1994a; Lee et al., 2006). The work on BCoV demonstrates the important role of cleavage sites in shifting coronavirus genetic populations depending on the proteolytic environment.

3.4.4. Infectious bronchitis virus (IBV)

The avian gamma-coronavirus IBV was the first described coronavirus. It provokes extensive impairment of respiratory and urogenital tracts of chickens, and has been well studied. Many strains of IBV have been characterized, most of which have a restricted cell and tissue tropism, often limited to primary chicken cells. All IBV strains harbor a well-defined furin cleavage site at the S1/S2 site (R-R-F-R-R-S, Table 1) that is cleaved during protein biosynthesis. Strikingly, a lab adapted strain of IBV, the Beaudette strain, which has been heavily egg- and cell-culture adapted has acquired a mutation at its S2* site, giving rise to the following motif S-R-R-K/R-S. Such a site is deemed ideal for furin recognition and cleavage as it harbors the critical P1 and P4 arginines (R), along with a highly favorable basic P2 lysine (K) or arginine (R) as well as flanking serines (S) at P1 and P6 (Yamada and Liu, 2009). Notably, the IBV-Beaudette S2* site is also cleaved during protein biosynthesis in S-transfected Huh-7 cells and in IBV-Beaudette-infected Vero cells, and this has been shown to be important for syncytiogenesis formation as well as entry of the virus (Belouzard et al., 2009; Yamada and Liu, 2009). Furthermore, IBV-Beaudette has a characteristically widened tissue and cell tropism and can productively infect many different cell types in vitro, including human cells. Importantly, high levels of furin expression, such as that observed in Huh-7 cells, was found to correlate tightly with highly productive infection by IBV-Beaudette produced in Vero cells (Tay et al., 2012). The wide cell and tissue tropism is in sharp contrast with the very restricted cell and tissue tropism observed for the other strains of IBV, which do not harbor a S2* furin cleavage site. Similarly to the concept found for avian influenza, modification of its S2* cleavage site may have allowed IBV-Beaudette to be cleaved by ubiquitously expressed furin, allowing for this expansion in tissue and cell tropism. This unusual feature is shared with MERS-CoV, as described in paragraph 2.4.7. MERS-CoV.

3.4.5. Porcine epidemic diarrhea virus (PEDV)

An alphacoronavirus, PEDV causes severe outbreaks of diarrheal disease in pigs mostly centering in East Asia (Park et al., 2011b). The virus recently emerged in the United States and is of particular concern because of the high mortality rates associated with piglet infections. In an early study, trypsin was found to activate PEDV for infection in Vero cells (Hofmann and Wyler, 1988). The S protein of Vero cell–produced PEDV was shown to be cleaved by trypsin when the virus is engaged with its receptor, but not on free virions (Park et al., 2011a). This suggests that PEDV S protein requires receptor-binding–induced conformational changes in order for trypsin cleavage to occur. TMPRSS2 was found to play a role in releasing PEDV particles from Vero producer cells stably expressing TMPRSS2 (Shirato et al., 2011). PEDV produced in Vero cells was also shown to enter cells via a clathrin-dependent endocytic pathway, also involving low pH and the activity of serine proteases (Park et al., 2014). In a recent study, notable findings about proteolytic activation of PEDV S have been uncovered (Wicht et al., 2014b). The authors focused on analyzing and comparing two strains of PEDV propagated in Vero cells, the first, CV777 which is an isolate that, like most PEDV isolates, requires trypsin activation for infection in cell culture, and the other, DR13, is a cell culture adapted strain, and has evolved to become trypsin-independent. The authors were able to demonstrate that CV777 S harbored a critical arginine (R) residue at the S2* site that is cleaved by trypsin (Table 2). Notably, they found that in the cell–culture adapted strain DR13, an R to G mutation occurred at the S2* site and explains why the strain has evolved to become trypsin-independent. As most coronavirus S proteins contain an arginine (R) at the S2* site (Table 2), it would be interesting to study whether the findings of Wicht et al. can be generalized.

3.4.6. Feline coronavirus (FCoV)

Studies on FCoV offer deep insights into the role of S and in particular its proteolytic processing and fusion activation in coronavirus pathogenesis and host cell tropism. The FCoV is an alphacoronavirus and there are two known serotypes, based on differential S antigenicity (Pedersen, 2009). While serotype II viruses are the most extensively studied, notably due to their propensity to grow well in vitro, such as in CrFK feline kidney cells, they are not the most clinically prevalent. The serotype I is the most prevalent but is difficult to culture and so has yet to be extensively studied, although some strains, such as Black (TN-406) and UCD1, can be grown in feline AK-D cells and feline macrophage-like cell Fcwf-4, respectively (Neuman et al., 2006; Pedersen, 2009). In both serotypes I and II, FCoV exists as two biotypes or pathotypes that display extremely contrasting pathogenicity. The low pathogenicity biotype, feline enteric coronavirus (FECoV) is a highly prevalent virus that transmits easily through the fecal–oral route and generally causes only mild and self-limiting enteritis. In contrast, the highly pathogenic biotype, feline infectious peritonitis or FIPV, causes an acute, systemic and immune-mediated disease that is invariably fatal and represents a leading cause of infectious...
disease deaths in cats. The current understanding on FIP pathogenesis, the so-called “internal mutation theory”, is that it arises from an initial FECV infection in which the virus acquires mutations within the infected animal that leads to its conversion from the low pathogenic form to the virulent FIPV form (Vennema et al., 1998). A key difference between FECV and FIPV is their tissue and cell tropism: while FECV is generally restricted to the epithelial cells of the gastrointestinal tract, FIPV has the ability to readily infect monocytes and macrophages (Stoddart and Scott, 1989). It is this change in cell and tissue tropism, from gut epithelial cells to motile immune cells, that is thought to be the crucial step toward systemic spread of the virus. Several studies have mapped mutations that are linked to the transition between the biotypes, located in the accessory 3c and ORF 7 genes (Herrewegh et al., 1995; Pedersen et al., 2012), but there is a growing body of evidence indicating that mutations within the S protein are key to the development of FIP (Chang et al., 2012; Licitra et al., 2013). In the following section we will describe the current understanding on the S mutations that are thought to be linked with FIP.

### 3.4.6.1. Type I feline coronavirus (FCoV)

Type I FCoV has S proteins harboring two putative cleavage sites: S1/S2 and S2’ (Tables 1 and 2). A study on serotype I strains of FCoV has demonstrated that alphacoronaviruses could harbor a cleaved S protein (de Haan et al., 2008). In particular, serotype I UCD strain, a FECV obtained from the feces of experimentally infected cats, was shown to contain a S1/S2 site (R-R-S-R-R-S) that is cleaved by furin. In contrast, the authors show that UCD1, a FIPV adapted for growth in Fcwf cells, has acquired a mutation at the P1 position (R-R-S-R-G-S) abrogating cleavage by furin. It was also shown that because the S1/S2 site in UCD1 S was left uncleaved, it retained an intact heparan sulfate binding motif (B-B-X-B; B: basic residue), allowing the virus to bind the polysaccharide, a property that is not shared with its parental strain UCD. Since UCD1 arose from cell-culture adaptation, it is an important example of how a change in S cleavage site can lead to acquisition of new characteristics, such as heparan sulfate binding, that can lead to modulation of cell tropism and pathogenicity.

Licitra and colleagues have studied S1/S2 cleavage site sequence diversity found in circulating field strains of type I FCoV and compared them with those found in FIPV (Licitra et al., 2013). Remarkably, FECV S sequences harbored a highly conserved furin cleavage motif (R-R-S/A-R-R-S), an unusual property for an alphacoronavirus, which is shared with type I canine coronavirus (Table 1). However, a strong correlation was found between FIPV S sequences and presence of one or more non-silent mutations at the S1/S2 cleavage site, introducing a range of hydrophobic or non-polar residues. Biochemical assays employing fluorogenic peptide mimetics of the mutated cleavage sites confirmed that the substitutions found in FIPV abrogated furin cleavage. A more recent study, using an extended sample set of type I FCoV, and which compares S1/S2, including the P6 position, and the S2’ cleavage sites of FECV and FIPV, confirmed the strong association between FIPV mutations and cleavage sites at the S1/S2 and S2’ cleavage sites (Licitra et al., manuscript in preparation). It is hypothesized that mutation at the S1/S2 and S2’ cleavage sites may switch S activating protease requirements, allowing for a change in tissue and cell tropism. Further research is required to investigate this hypothesis. A practical outcome of this discovery would be to use the type I S1/S2 and S2’ sites as biotype markers to discriminate between type I FECV and FIPV infections.

### 3.4.6.2. Type II feline coronavirus (FCoV)

The type II FCoV has a S protein that is distinct from type I FCoV as it was demonstrated to have arisen from recombination with canine coronavirus (CCoV) S (Herrewegh et al., 1998). Although the precise mechanisms involved differ from type I FCoV, a study on the prototypical FECV (1683) and FIPV (1146) type II FCoV strains, produced in canine A-72 cells, has shown that there is also a link between modified proteolytic processing requirements and cell tropism (Regan et al., 2008). Whereas FECV-1683 entry was shown to be highly dependent on endosomal cathepsin B and L activity and low pH, FIPV-1146 was shown to be dependent on cathepsin B activity but not cathepsin L activity. Importantly, while FECV-1683 S was demonstrated to be cleaved by both cathepsin B and L, FIPV-1146 could only be cleaved by cathepsin B. This difference in proteolytic processing may influence which cell types each virus infects, as individual protease expression profiles differ from cell type to cell type. It would be important to investigate further whether this change in proteolytic activation requirement is due to a mutation found within S, particularly at the S2’ site of FIPV-1146 S, as the R to G mutation found at that site (Table 2) is reminiscent of the R to G mutation found in PEDV S2’ (Wicht et al., 2014b).

### 3.4.7. Middle East respiratory syndrome coronavirus (MERS-CoV)

MERS-CoV, a recently emerged virus from the Middle East that is associated with severe pneumonia with a high case fatality rate, is another example of a coronavirus whose S protein can be activated by a multitude of proteases. Since its isolation and discovery in 2012, the complete viral genome was swiftly sequenced (van Boheemen et al., 2012), and its receptor, dipeptidyl peptidase 4 (DPP4 or CD26), rapidly identified (Raj et al., 2013).

In a report using pseudovirions produced in HEK-293T cells as well as protease inhibitors, it was first shown that TMPRSS2 and endosomal cathepsins could activate MERS-CoV virus–cell fusion (Gieren et al., 2013). The authors also reported the sensitivity of MERS-CoV entry to the lysosomotropic weak base NH4Cl, a result attributed to a decreasing of activity of endosomal cathepsins, which require low pH.

The role of cell surface-expressed TMPRSS2 in the activation of MERS-CoV was confirmed using native MERS-CoV virions produced in Vero cells, and it was estimated that the entry of the virus in Vero cells overexpressing the protease was increased 100-fold, compared to control cells (Shirato et al., 2013). The authors described how TMPRSS2 could greatly potentiate MERS-CoV infection-induced cell–cell fusion in Vero cells stably expressing the protease. The study also corroborated that endosomal cathepsin L could activate MERS-CoV S. Notably, the study showed that depending on the target cell type, MERS-CoV could use different activation pathways.
4. Discussion: Spike (S) cleavage motifs as coronavirus cell and tissue tropism, host range, and virulence markers

In recent years, it has become more appreciated that the protease-mediated activation of membrane fusion is a versatile process that can be exploited by coronaviruses not only for the basic requirement of entry into host cells, but also to modulate this entry process and allow changes in cell, tissue and species tropism. In concert with changes in the receptor binding site(s) of S, the ability to modulate infectivity via cleavage site changes makes coronaviruses of particular concern as emerging pathogens.

In particular, a recent study on MERS-CoV has strengthened the concept that along with the virus receptor, the repertoire of proteases a given cell type expresses can be a determining factor for infectivity (Barlan et al., 2014).

Our present understanding is that probably all coronavirus S are cleaved at some point during infection, in line with their designation as class I viral fusion proteins. In many cases, cleavage occurs at the S1/S2 position; a location that (based on the situation with other class I viral fusion proteins such as influenza HA) is at the expected junction between the S1 (receptor binding) and S2 (fusion) domains. We believe that in most cases, this cleavage will occur during the process of S protein maturation and virus assembly. S1/S2 cleavage may not be obligatory, however. While in several cases, coronaviruses appear to be released from cells with uncleaved S in many other cases the S1/S2 site is cleaved, often by furin. Previously, it has been suggested that alphacoronaviruses have uncleaved S, butacoronavirus may or not be cleaved, with gammacoronaviruses having obligatory cleavage. However this now appears to be an oversimplification; for instance, S proteins of alphacoronaviruses in the feline and canine serotype 1 lineage have strong furin cleavage sites and have been shown to be furin cleaved, and based on bioinformatics (Table 1), S proteins of gammacoronaviruses in aquatic mammalian hosts appear not to be cleaved by furin. It appears that different coronaviruses have different requirements for S1/S2 cleavage.

The more recently identified cleavage site within the fusion domain (S2′) is more directly connected to the proposed fusion peptide, and so may function in a more analogous manner to the influenza HA cleavage site. As such, we believe that S2′ cleavage is an obligatory event for coronavirus entry (Burkard et al., 2014), with the possibility of cleavage at the S1/S2 site priming subsequent cleavage at the second (S2′) site (Belouzard et al., 2009; Millet and Whittaker, 2014). There is accumulating evidence that S2′ cleavage mainly occurs during virus entry (either at the cell surface or in endosomal compartments), and may be a relatively transient event. The transient nature of the cleavage means that in many cases it is difficult to observe this cleavage event based on conventional techniques (e.g. Western blots). While the nature of the proteases cleaving at S2′ remain unclear, it is likely that cathepsins would be involved, as well as potentially cell-surface TMPRSS2-like proteases. Bioinformatic analysis of S2′ indicates that a strong furin motif is present on laboratory-selected viruses (IVB-Beaudette and BCoV-B2.78.B0.P1, Tables 2 and 3). The presence of a robust furin cleavage site on MERS-CoV at S2′ may represent a novel, if not unique, situation for a naturally occurring coronavirus. However, bioinformatics also predicts that some alphacoronaviruses (PEDV, TGEV, type II FCoV and type II CCoV) also have pronounced furin cleavage motifs at S2′. It will be important to test this possibility experimentally.

Another parallel to point out between cleavage sites of coronaviruses and influenza viruses is that it appears that the presence of a furin cleavage site adjacent to the putative fusion peptide can arise by either insertion of a polybasic site, as is the case for HPAI viruses of the H5 and H7 subtypes (Kawaoka and Webster, 1988; Lee et al., 2006) and the bovine coronavirus B2.78.B0.P1 (Table 3) (Borucki et al., 2013), or by incremental substitutions such as in avian influenza virus H9 subtype (Tse et al., 2014), MERS-CoV (Millet and Whittaker, 2014), or IVB-Beaudette (Tay et al., 2012,Table 2).

Notably, MHV S contains what appears to be an insertion of a stretch of residues upstream of the cleavage site (Table 3), which are clearly linked to fusion activity (Taguchi and Shimazaki, 2000). In this regard, there are also parallels with influenza virus, where peptide insertions upstream of the cleavage site/fusion peptide modulate the fusion activity of H7 influenza viruses (Hamilton et al., 2012; Perdue, 2008). In addition, work on influenza virus HA have shown that glycosylations can have profound effects on cleavage activation, as bulky sugar moieties can create steric hindrance around the cleavage site, restricting access for proteases (Kawaoka et al., 1984; Tse et al., 2014). The role of glycosylations in modulating prolytic activation of the coronavirus S protein, which is known to be heavily glycosylated, has been relatively unexplored and awaits further investigation.

Inhibitors of host cell proteases, including TMPRSS2, PCs and furin, have been proposed as potential antiviral therapeutics for coronavirus infections, in particular for the treatment of SARS and MERS infections (Bergeron et al., 2005; Gierer et al., 2013; Millet and Whittaker, 2014; Simmons et al., 2013). Targeting host cell proteases in the context of a virus infection presents the advantage of minimizing the development of viral drug resistance observed when targeting virus proteins. The list of cellular protease inhibitors approved for clinical use for various diseases, such as trypsin-like, factor X, and neutrophil elastase inhibitors, is growing (Turk, 2006). However, it is important to consider that such treatments increases the risk of higher toxicity and side effects as host cell proteases, such as furin, are essential for a great number of cellular processes (Seidah and Prat, 2012; Turk, 2006). Clearly, targeting host cell proteases will require more research efforts before actual therapeutics can be safely and effectively used against coronavirus infections.

Many open questions remain, including the specific proteases involved, the compartment and pH at which cleavage occurs, what proportion of S need to get cleaved within a single virus particle/S trimmer and how many virus particle in a population get cleaved. New technologies such as single particle studies and selective biotinylation assays will hopefully provide answers to these questions in the future (Costello et al., 2013; Wicht et al., 2014a). In addition to the S1/S2 and S2′ sites, it is also possible that the S protein gets cleaved in other locations, giving rise to the concept that S undergoes a progressive “destabilization”, based on a combination
of proteolysis (sometimes at multiple sites), low pH, and receptor binding, as part of its fusion activation (Gierer et al., 2014; Simmons et al., 2011). In this regard coronaviruses may show more similarity to Ebola virus, another virus with a class I fusion protein, than it does to influenza virus (White et al., 2008). It is also important to note that many coronavirus receptors are membrane ectopeptidases (ACE2, APN, DPP4). To date, however, there is no evidence for proteolytic involvement of these receptors in virus entry (Bosch et al., 2014).

Bioinformatic analysis of the coronavirus S cleavage sites suggests the possibility of using this data within the framework of virulence or bio-markers that can differentiate different coronaviruses. For type I FCoV S, the S1/S2 and S2 cleavage sites are likely to be a biotype markers differentiating FECV from FIPV, and for type II FCoV S a similar situation may also occur for their S2 site. Notably, the amino acid sequence within the two cleavage sites are often highly conserved within groups of viruses, and so sequencing information of the cleavage sites may be useful in the future for virus typing, rapid identification of new viruses and as a “virulence marker”. This proposed virus typing can be applied to so-called polytropic coronaviruses, in particular IBV-Beaudette and MERS-CoV. These studies indicate that a second furin cleavage positioned at the S2 site may be viewed as a marker of expanded cell and tissue tropism, host range and perhaps pathogenicity (Millet and Whittaker, 2014; Tay et al., 2012). At present, modulation of the S2 site seems to be correlated most strongly to changes in tissue and cell tropism, host range and pathogenesis (e.g. IBV, MERS-CoV, PEDV). However, in other cases changes in the S1/S2 site are also strongly correlated (e.g. type I FCoV) with modification of cell tropism and pathogenesis. Overall, it seems likely that modulation of either of two protease cleavage sites by coronaviruses can have a profound impact on disease outcome, depending on the individual coronavirus.

The coronavirus field has gained tremendously thanks to the study of S proteolytic processing by host cell proteases. As shown in this review, such investigations are crucial for the development of the study of coronavirus entry, cell and tissue tropism, host range, and pathogenesis. While many questions remain, we look forward to future breakthroughs and discoveries in this active and lively area of investigation.

Acknowledgements

We thank Weishan Huang for critical reading of this manuscript.

This work was supported by National Institutes of Health Grant R21 AI110805.

References


Pu, Raj, Park, Park, VIRUS Peng, Shirato, G Model

Please note that the text has been automatically converted to natural language. The document includes references to various topics related to virology, including SARS-CoV, coronavirus, human infection, and viral mechanisms. It also refers to specific individuals and works, such as Dr. Raj, Park, and various researchers mentioned in the footnotes. The text covers a range of topics, including the role of viral entry mechanisms, the importance of viral receptors, and the implications of viral pathogenesis.

The document appears to be a discussion or review article, possibly related to the impact of viral infections on human health and disease. The references and data points indicate a detailed analysis of viral behavior, including the role of viral spikes, cell receptors, and the implications of viral entry into host cells.

Please cite this article in press as: Millet, J.K., Whittaker, G.R., Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. Virus Res. (2014). http://dx.doi.org/10.1016/j.virusres.2014.11.021