Virus Research 202 (2015) 120–134

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis

Jean Kaoru Millet, Gary R. Whittaker*

Department of Microbiology and Immunology, Cornell University, C4 127 VMC, Ithaca, NY 14853, United States

ARTICLE INFO

Article history:
Available online 22 November 2014

Keywords:
Cleavage activation
Coronavirus
Pathogenesis
Protease
Spike protein
Tropism

ABSTRACT

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 between 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 fusion 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.

1. Introduction

Coronaviruses are a wide-ranging family of viruses that infect many species of birds and mammals, including humans (Woo et al., 2009). They possess a remarkable ability for interspecies transmission as exemplified by the emergence of the deadly human virus severe acute respiratory syndrome coronavirus (SARS-CoV) (Drosten et al., 2003; Peiris et al., 2003, 2004), and more recently, Middle East respiratory syndrome coronavirus (MERS-CoV) (van Boheemen et al., 2012; Zaki et al., 2012), both of which are thought to have originated in bats (Li et al., 2005b; Wang et al., 2014), followed by an intermediate host stage (civet cats and camels, respectively) (Alagaili et al., 2014; Haagmans et al., 2013; Hemida et al., 2014; Wang and Eaton, 2007), before crossing into the human population (Drexler et al., 2014). Such zoonotic potential is of particular concern, especially since global trade, deforestation, massive urbanization and high density farming practices increase the likelihood of sparking new and severe zoonotic outbreaks (Cutler et al., 2010).

The success of coronaviruses in their ability to jump between species may be attributed, in part, to the diverse array of virus entry strategies they deploy to infect target cells (Belouzard et al., 2012; Bosch and Rottier, 2008). Coronavirus entry is largely controlled by the spike surface envelope glycoprotein (S) since it bears both receptor binding and membrane fusion capabilities (Masters and Perlman, 2013). As such, the S glycoprotein is a crucial determinant of tissue and cell tropism as well as host range. Coronaviruses are notable because at each step of virus entry, which includes receptor binding, activation of fusion, and internalization, a multitude of mechanisms and strategies have evolved (Belouzard et al., 2012). For example, depending on the coronavirus species, the S protein can mediate binding to a proteinaceous receptor or to carbohydrate moieties.

Coronaviruses can enter cells via fusion either directly at the cell surface or can be internalized through the endosomal compartment. The mouse hepatitis virus (MHV) is a prime example of the flexibility in entry mechanisms used. A variant of the MHV-4 strain was shown to be able to fuse directly at the cell surface at

* Corresponding author. Tel.: +1 607 253 4019.
E-mail address: grw7@cornell.edu (G.R. Whittaker).

http://dx.doi.org/10.1016/j.virusres.2014.11.021
0168-1702/

neutral pH and 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 spatio-temporally 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 characteristically 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 (HPAI) 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, HA_1 which binds to sialic acid receptors and HA_2 which contains the fusion machinery, featuring mainly alpha-helical secondary structures (Skehel and Wiley, 2000; White et al., 2008). HA_2 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 HA_0. 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 HA_0 into HA_1 and HA_2 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 HA_2 buried at the subunit interface, and is exposed upon cleavage of HA_0 and conformational changes of HA.

A key initial step in the HA fusion process is its priming by proteolytic cleavage that separates HA into HA_1 and HA_2 fragments. At this stage, the HA is in a metastable "spring-loaded" conformation. The receptor-binding subunit HA_1 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 HA_2 (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