Infectious Bronchitis Virus (Gammacoronavirus) in Poultry: Genomic Architecture, Post-Translational Modifications, and Structural Motifs


1. Introduction

Infectious bronchitis virus (IBV) is an avian coronavirus (CoV) that belongs to the genus Gammacoronavirus and has been listed as an important disease by the World Organization for Animal Health (WOAH). It causes highly contagious respiratory, reproductive, and renal diseases in commercial poultry farms. Multiple IBV serotypes and genotypes have been identified in many countries and many detected variants do not provide cross-protection against infection, resulting in repeated outbreaks and significant economic losses worldwide. In addition, the high genetic mutations and recombination events in the prominent genomic regions of IBV, particularly in the spike glycoprotein (S) and nucleocapsid (N) proteins, are directly involved in the evolutionary processes of IBV and lead to increased pathogenicity and tissue tropism. The characterization of the different genotypes and the relationship between the structure, function, post-translational modifications (PTMs), and structural motifs will elucidate the mechanisms that promote replication and pathogenicity and affect the host’s immune response during infection. In this review, we discuss the molecular features of various IBV genes and proteins that contribute to the infection process. We also highlight the common PTMs and structural motifs that occur during protein synthesis and are essential components of IBV ecology.

Keywords: IBV; Gammacoronavirus; molecular structure; motif
can occur in broiler or breeder chickens as a consequence of secondary infection with *Escherichia coli* and Mycoplasmosis [8,9]. Moreover, the serious economic consequences are also associated with high production losses, retarded growth, poor health, and high rates of condemnation of meat in broiler processing. IBV disease has been considered an epidemic that might be commonly spread in both vaccinated and non-vaccinated poultry farms [10]. The outbreaks of IBV are quite intermittent due to the frequent emergence of strains or variants with high mutation and replication rates in the hypervariable region (HVR) of the S gene [11]. According to Legnardi et al. [12], the S1-based hypervariable segment has been sequenced to organize IBV variants into genotypes for a long period of time due to strain heterogeneity and linkage with genetic topology. Studies have discovered more than 50 distinct IBV antigenic variants using virus-neutralization assays and molecular analysis of the S1 protein gene [13]. Moreover, these variations exhibit a limited capacity for cross-protection against other strains. In addition, the IBV genome undergoes frequent recombination events which lead to emerging chimeric viruses as new genotypes [14,15].

The global poultry–meat trading industry and the movement of farm personnel and migratory birds are major factors contributing to the global transmission of IBV, leading to extensive ecological diversity and substantial variations in IBV strains [16]. Although inactivated and live attenuated vaccines have been employed to prevent or control IBV, outbreaks are still persistent [17]. This suggests that the existing vaccines may not provide complete protection against the diverse array of IBV strains due to their rapid evolution and recombination. The continuous emergence of new IBV variants further emphasizes the need for effective surveillance, vaccine development, and control strategies to combat the ongoing challenges posed by IBV outbreaks. Despite efforts to control IBV infection through vaccination, complete elimination has not been achieved. One of the reasons for this is the inability of vaccines to induce cross-protection against the wide range of genotypes, variants, and serotypes that emerge in the field [18,19]. According to a study conducted by José et al. [20], it was found that an unidentified avian CoV serves as a donor for the structural and accessory genes of the field strain. The continuous recombination between the unidentified avian CoV and IBV vaccine strains highlights the dynamic nature of IBV evolution and the potential for genetic exchange between different viral strains [21]. Therefore, it is necessary to identify the potential specific IBV genotypes or strains to control particular geographic regions and develop new vaccination strategies against the disease [22]. It was proven that prime-boost programs combining both classically and genetically related variant IBV vaccine strains are essential for broad protection against different IBV strains during field challenges [23]. The knowledge of IBV surveillance must place a greater emphasis not only on identification but also on genomic analysis in order to detect the emergence of novel and dominant lineages [24]. Therefore, it is essential to highlight the significance of IBV heterogeneous genomes with different IBV emerging strains or variants to establish the appropriate protocols for vaccination. The current review discusses the genomic structure of IBV and post-translational modification (PTM) during pathogenesis as well as the impact of genetic diversity on the emergence of IBV evolution.

2. Genomic Structure of IBV

IBV belongs to Gamma or Group-3 CoV (order Nidovirales, family Coronaviridae) with a positive-sense single-stranded RNA (+ssRNA) virus and a diameter of 120–160 nm. The virions of IBV are enclosed in spike-like projections approximately 20 nm in diameter [3]. According to antigenic cross-reactivity and nucleotide sequence analysis, CoVs are classified into four groups: Alpha, Beta, Gamma, and Delta [25]. Alpha, or Group-1, CoVs include the human CoV, porcine transmissible gastroenteritis CoV (TGEV), and porcine respiratory CoV (PRCV); Beta or Group-2 CoV include severe acute respiratory syndrome CoV (SARS-CoV), COVID-19, MERS-CoV, mouse hepatitis CoV (MHV), and bovine CoV (BCoV) [26,27]; and Gamma or Group-3 and Group-4 CoV are generally associated with all kinds of fowl, especially chickens, pheasants, and Galliformes [28]. Historical evidence suggests that interspecies recombination has occurred in avian CoVs, as observed in the
remarkable similarity between the genome of turkey CoV (TCoV) and IBV, except for the S gene [29]. These interspecies recombination mechanisms have also been reported in other Coronaviridae family members, such as the SARS-like CoV from bats, which may have developed via recombination between another SARS-like CoV (Bt-SLCoV) and a human SARS-CoV [30].

The length of the IBV genome is approximately 27.6 Kb and the virion is encircled by 5′ and 3′ untranslated regions (UTRs) with a poly (A) tail [8,31] along with their functional and nonfunctional genes as shown in Figure 1. Nine functional genes have been identified in the IBV genome, four of which encode main structural proteins and another five genes that encode nonstructural proteins (Nsps) [32]. All CoVs, including IBV, contain an exoribonuclease (ExoN) protein related to the RNA-dependent RNA polymerase (RdRp) that is responsible for editing or proofreading functionality during replication [33]. The main portion of the genome is covered with two overlapping open reading frames (ORFs) that can be translated into large polyproteins 1a and 1b via a ribosomal frameshift mechanism [34]. The coding regions cover the remaining part of the genome and the leading structures of proteins are the spike (S), envelope (E), membrane (M), and nucleocapsid (N). However, two additional genes, ORF3 and ORF5, encode the accessory proteins 3a and 3b plus 5a and 5b, respectively [17,35]. Li et al. [36] have reported on its new genome enclosed by two variants of transcription-regulating sequences (TRSs) that exhibit a highly conserved TRS: CTTAACAA, positioned in front of ORFs 1a, 1b, 3a, M, 5a, and N, and a variant, AAGAACAA, in front of the S gene.

![Figure 1. Morphology and genomic structure of IBV with different structural and nonstructural viral genes (adapted from Bhuiyan et al.) [37].](image)

The polymerase gene consists of ORF1a and ORF1b, which encompasses approximately two-thirds of the genome at the 5′ ends. These genes are encoded in the replicase transcriptase proteins and overlaid in ORFs that are highly conserved. These proteins facilitate ribosomal frameshifting mechanisms through transcription, replication, translation, and packaging of the viral genome complex [38]. The ribosome alters the ORF during transcription in response to the presence of the heptanucleotide “slippery sequence” UUUAAC which is positioned upstream of a pseudoknot structure in the RNA [39]. Ribosomal frameshifts allow the translation of both 1a and 1b ORFs as a large polypeptide. These polyproteins are post-translationally cleaved to produce approximately 15 Nsps (Nsp 2–16). The main protease is collectively known as the 3C-like protease, a papain-like protease (PLP), which constitutes the replication complex termed the RNA-dependent RNA-polymerase (RdRp) [40]. As a group, there is no evidence that the genetic signature for Nsp 1 has been encoded in the genome of avian CoV (Type III), as it is not necessary for CoV RNA synthesis. However, there are only 15 Nsps in IBV while genomes in other CoV groups encode 16 Nsps [41,42]. The particular role and mechanism of every nonstructural
protein inside the replication process have not been elucidated but they occasionally act as putative enzymes and functional domains. The functional activity of proteins is essential for the CoV replication complex, which comprises various NspS such as Nsp 3 (PLP) [43], Nsp 4 and Nsp 6 (membrane-bound proteins) [44], Nsp 8 (RdRp activity) [27], Nsp 7 to Nsp 10 (RNA binding activity) [45], Nsp 12 (RdRp) [46], Nsp 13 (RNA helicase) [47], Nsp 14 (3’-5’ exonuclease) [28], Nsp 15 (endoribonuclease) [48], and Nsp 16 (methyltransferase activity) [49].

Gene 2 consists of a single ORF that encodes the spike glycoprotein (S). This protein consists of four domains anchored in the lipid bilayer of the virion and is post-translationally cleaved into the subunits of amino-terminal S1 (92 kDa) and carboxy-terminal S2 (84 kDa) [50]. The furin-mediated cleavage of the S protein normally occurs at the locations of RRRR537/S and RRRR690/S [51,52]. The spike glycoprotein is a comprehensively glycosylated membrane protein with three structural domains [53,54]. In the mature virion, the S2 portion is constantly associated with S1 and the protein–membrane anchorage of S1 forms the multimeric coiled-coil S protein [7]. The S1 subunit is located on the surface of the virion and is responsible for the fusion of the viral envelope and cell membrane of the host [55,56]. Although the primary receptor for IBV has not been reported, other CoVs have used cellular receptors such as neuraminic acid, sialic acid, aminopeptidase, carcinoembryonic antigen, or angiotensin-converting enzyme-2 (ACE-2) [57,58]. The specific molecular interactions between the viral S1 protein and the cellular receptor binding site dictate the efficiency of viral entry into host cells. However, animal CoVs show reduced affinity for binding to the human ACE-2 receptor, which is believed to be a significant barrier for animal CoVs to establish efficient infections in humans [59]. The S1 portion of the S gene plays an important role in viral attachment through membrane fusion and binding activity via sialic acid receptors, which is considered a crucial factor for viral diversity and escape from immune defense [60]. Additionally, the S1 subunit contains virus-neutralizing and hemagglutination-inhibiting antibodies which play a crucial role in the recognition and binding of specific amino acids for identification and interaction with host cells [61]. In the most recent categorization of IBV, which is based on S1 sequencing, 32 lineages and 6 genotypes were established with pairwise genetic distances of 30% and 13%, respectively [12]. The majority of IBV lineages are restricted to particular geographical areas; some countries have distinct lineages. The GI-1 and GI-13 lineages, also known as Mass and 793B kinds, are frequently found in different nations because of the usage of vaccines produced from these lineages, while the GI-1, GI-13, GI-16, and GI-19 lineages are extensively dispersed [11,62]. Recently in China, researchers discovered two new types of IBV through genetic analysis of various IBV isolates known as GI-28 and GI-29, which are distinct from the previously identified 32 lineages and recombinant variants. The existence of such a diverse range of lineages and genotypes poses a significant challenge for commercial vaccines, even vaccinated chicken flocks are susceptible to new outbreaks of IBV due to this extensive genetic diversity.

The S1 sequences from different strains vary significantly between 2–25% at the amino level, while the S2 subunit is conserved in viral attachment through membrane fusion [63]. Likewise, the ectodomain region of the S2 subunit is covered with a fusion peptide-like region and two heptad repeats located at 100–130 Å in length around 771–879 amino acids in the IBV genome. The S2 portions are incorporated in the oligomerization of the protein which helps to access susceptible host cells. [63,64]. However, the latest research in CoV biology has proposed that the S2 subunit of the S protein could potentially influence the cell tropism of the virus. In further studies, the lack of the S2’ cleavage motif was shown to result in the loss of the ability of the Beaudette strain to replicate in Vero cells [65]. The high variation of the S gene nucleotide sequences might cause less cross-protection between serotypes and alter the immune capability of a vaccine [3]. However, the S glycoprotein has targeted genetic characteristics such as the genotype, serotype, and pathotype classifications which can help produce recombinant vaccines for IBV [66].
Gene 3 consists of three ORFs designated as 3a, 3b, and 3c where ORF 3c encodes the envelope protein (E). ORF 3a and ORF 3b are enclosed within highly conserved nucleotide sequence regions in the IBV genome [67]. The characterization of the 3a and 3b proteins has not been studied so far. Nevertheless, there is a hypothesis suggesting that both proteins may not be essential for viral replication in vitro for IBV and MERS-CoV. On the contrary, separate studies conducted on IBV have shown that removing either the 3a or 3b protein can result in attenuation of the virus both in vivo and in vitro [65]. The IBV-E protein (ORF 3c) is a small polypeptide of approximately 9–12 KDa. This covers only a minor constituent of virions and contains a small hydrophilic amino terminus (8–12 residues) followed by a large hydrophobic region (21–29 residues), including 2–4 cysteines and a hydrophilic carboxy-terminal tail (39–76 residues). The viral membrane E protein was exposed to MHV and IBV through the principle of resistance to alkaline extraction [68]; membrane insertion occurs without cleavage of a signal sequence [69]. The IBV-E protein is highly necessary for viral assembly, budding, and ion channel activity but occasionally mutations occur in the amino acid sequence that causes significant interruption to the viruses assembling in cells [70,71]. It has been suggested that the IBV-E is a poorly characterized minor envelope protein that prevails in small amounts in virions; it is erroneously expressed at higher levels in infected cells, near viral assembly [72]. CoV-E stimulates apoptosis as recognized in MHV- and SARS-CoV-infected cells [70,73]. Westerbeck et al. [74] reported that IBV infection is stimulated to increase the Golgi luminal pH after the effects of IBV-E overexpression during infection, resulting in the N protein transforming the Golgi pH to shield the S protein and encouraging the release of the infective virus. Studies conducted on IBV have revealed that mutant viruses with changed E proteins replicated more slowly than wild-type viruses; electron microscopy also revealed malformed virions [64]. Although the diverse actions of the CoV-E protein are exhibited by different CoVs, the reason for this is not yet fully understood and requires further exploration [70].

Gene 4 consists of a single ORF encoding the M glycoprotein (27–36 KDa) which is exposed mostly in the exterior part of the virion (10% at the outer surface of the virus) or intracellularly in the lumen of the endoplasmic reticulum (ER) [75]. The M protein is a triple-spanning transmembrane (TM) protein enclosed by a minor portion of the ectodomain and a very large portion of the carboxy-terminal endodomain [76]. The main function of the M protein is to act as a center performer in the virion assembly of the virus particle by communicating with viral ribonucleocapsids and spike glycoproteins [77,78]. Previous studies reported that the M proteins of Group-1 and Group-3 CoVs are exposed to complete N-linked glycosylation [79]. The role of the glycosylation process for M proteins is not fully documented but it is capable of organ tropism in vivo and the ability to stimulate interferon (IFN) alpha in vitro [80]. Some studies suggested that the endodomain is the locus of interactions between the M protein and N protein [81,82] and the M protein and S protein [83], which can provide lodging for mutations or radical modifications [78].

Gene 5 contains two ORFs (ORF 5a and ORF 5b) that encode the 5a and 5b proteins, respectively, found in all three groups of CoVs [31,84]. Both proteins are functionally bicistronic and nonessential for replication in vitro using reverse genetics, as previously reported [85–87].

Gene 6 contains only a single ORF that encodes the N protein and ranges from 50–60 KDa [80]. The IBV-N protein is highly conserved and is associated with viral RNA, forming the helical ribonucleoprotein (RNP) complex [88]. The N protein is supported by three conserved domains, including one domain that interacts with RNA. The other two distinct domains are the amino-terminal RNA-binding domain (ATD) and the carboxy-terminal RNA-binding domain (CTD), separated by a flexible linker [77,89]. The IBV-N gene is present in the cytoplasmic region exposed to the nucleoli of approximately 10% of cells and it is closely associated with the replication and transcription processes [90]. Studies have gained valuable insights into the cellular localization of the IBV-E protein by utilizing antibodies that possess the ability to selectively bind to specific epitopes. Previous studies have revealed that the C-terminal region of IBV-E is predominantly found within
the cytoplasm of the host cell. In contrast, the N-terminal region of IBV-E undergoes translocation to a different cellular compartment. This information suggests that different regions of the IBV-E protein play distinct roles within the host cell and are involved in interacting with cellular factors or participating in intracellular processes specific to the replication and propagation of the virus. On the other hand, the translocation of the N-terminal region implies its involvement in interactions with components of another cellular compartment, potentially modulating host cell functions to favor viral replication or evasion of the immune system by interaction with genomic RNA to form a ribonucleoprotein (RNP) complex [66,91]. During transcription, the N protein is transferred from the nucleolus of the host cell to the cytoplasm as a policy to regulate the host environment and increase the concentration of host ribosomes during the interphase portion of the cell cycle [90]. Similarly, the basic functional residues of ATD are highly conserved, surrounded by CoVs of different antigenic groups [35,92]. Due to its high degree of conserved nature, targeting the N protein offers the potential to develop anti-IBV medicine and vaccines that can provide broad-spectrum protection against multiple strains of IBV and other related CoVs [93,94].

Nsps are the cleavage products of polyproteins 1a and 1b of Gammacoronavirus which are the main apparatuses of protein synthesis, especially RNA-binding activity and interaction with its RdRp [95–97]. Due to the cleavage site of the IBV replicase, only 15 Nsps (Nsp2–16) are produced without Nsp1. However, in Alpha and Beta CoVs, Nsp1 is generally located between Nsp1 and Nsp2 in the absence of cleavage [98]. According to sequence analysis and structural findings, the enzymatic activities essential for vital processes in CoV RNA synthesis have been mapped to some of the Nsps [35]. Tan et al. [95] reported that Nsp2, nsp5, Nsp8, Nsp9, and Nsp10 were found to bind to UTRs, although Nsp8 collaborates with both the N- and C-terminal regions of Nsp12. Conversely, Nsp8 has been adapted to interact with Nsp7 and act as a primase, synthesizing RNA primers for Nsp12. The differential activities of the Nsp proteins are provided in Table 1:

<table>
<thead>
<tr>
<th>Protein</th>
<th>Function</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nsp1</td>
<td>Amino-terminal protein lacking in IBV but existing in other CoVs function is the inhibition of cellular mechanisms, including translation and IFN signaling</td>
<td>[99]</td>
</tr>
<tr>
<td>Nsp2</td>
<td>Plays a vital role in assisting IBV protein synthesis by blocking protein kinase phosphorylation of eukaryotic initiation factor 2 (eIF-2alpha), which shuts down protein synthesis</td>
<td>[100]</td>
</tr>
<tr>
<td>Nsp3</td>
<td>Involves several domains comprising an acidic domain, an ADP-ribose 1 phosphatase, the PLP PLP1, and the TM domain; nevertheless, SARS-CoV is orthologous to PLP2 of other CoVs</td>
<td>[101]</td>
</tr>
<tr>
<td>Nsp4</td>
<td>A membrane-spanning protein, along with Nsp2 and Nsp6 is assumed to anchor the viral replication complex in double-membrane vesicles at the Golgi apparatus</td>
<td>[102]</td>
</tr>
<tr>
<td>Nsp5</td>
<td>An enclosing Mpro (cysteine protease) with a Cys–His catalytic dyad and responsible for cleaving the Nsps 4–16</td>
<td>[103]</td>
</tr>
<tr>
<td>Nsp6–10</td>
<td>Membrane-localized proteins that form a complex exhibiting replicase activity in the existence of an RNA primer</td>
<td>[104,105]</td>
</tr>
<tr>
<td>Nsp11–12</td>
<td>RdRp for viral RNA replication</td>
<td>[106]</td>
</tr>
<tr>
<td>Nsp13</td>
<td>RNA helicase with a function in unwinding or annealing RNA molecules</td>
<td>[107]</td>
</tr>
<tr>
<td>Nsp14</td>
<td>ExoN domain that provides support to RNA synthesis, proofreading, and repair</td>
<td>[108]</td>
</tr>
<tr>
<td>Nsp15</td>
<td>Endoribonuclease domain stimulated by retinoblastoma protein (pRb) in vitro</td>
<td>[109]</td>
</tr>
<tr>
<td>Nsp16</td>
<td>Methyltransferase and RNA cap formation</td>
<td>[110,111]</td>
</tr>
</tbody>
</table>

3. Diversity of IBV

According to evolutionary theories, genetic diversity and selection processes are the basic steps in the evolution of IBV [112]. The analysis of viral evolution relies on the investigation of nucleotide differences in the RNA of viruses which serve as fundamental evidence
for understanding evolutionary processes. Several factors contribute to the favorable conditions for IBV evolution, including the genome structure of CoVs (large positive-sense RNA), viral ecology (limited proofreading activity of viral polymerase), current poultry farming practices, and immunological pressure resulting from the widespread use of live vaccines [113,114]. The variation occurs in different strains due to mutations, insertions, deletions, and/or recombination [115,116]. The combination of mutations and recombination can allow the virus to move to the host cell and cause tissue tropism. Quinteros et al. [20,39] hypothesized that the unknown CoV donor of the structural and accessory genes of the field strain is continuing to recombine with IBV vaccine strains, leading to the emergence of novel strains.

The comparison of IBV gene sequences followed by their phylogenetic relationships helps in understanding disease emergence and epidemiology. Based on the IBV genome, structural genes such as the spike gene (S1, S2), M gene, and N gene were selected for sequencing and for conducting comparative studies among various IBV strains [117–119]. The alteration of nucleotide sequences is utilized to generate antigenic diversity within the genomes of different IBVs [8,120]. The S1 sequences from different strains exhibit significant variation at the amino acid level, ranging from 20% to 25% [121]. For example, the Connecticut and Mass strains, despite having different serotypes, showed a variation of only 7.6% in their S1 proteins at the amino acid level [16]. Similarly, sequence analysis and VN tests have demonstrated that certain isolates share more than 97% identity with the D274 strain, yet exhibit different serotypes. The position of a specific strain in a phylogenetic tree may vary depending on the genotyping techniques employed or the specific region of the genome being analyzed [122]. In certain instances, sequencing of the S1 and N genes was conducted to facilitate the exploration of tissue tropism in different IBV strains [123].

A phylogenetic analysis of 793/B strains using pairwise evaluation revealed that the 5′ terminus of the S1 gene exhibited a high amino acid sequence identity of 88.2% with the 4/91 strain and 86% with the H120 strain in the 3′ terminal sequence [124,125]. IBV is known for its high genetic diversity; numerous genetic lineages were identified based on the sequence of the S1 protein, which contains the HVR3 region. The main factor behind the shared clustering of the HVR3 sequences in these two or more lineages is the occurrence of various genetic alterations that have played a significant role in distinct clustering separate from other sublineages or co-circulation of divergent lineages belonging to different genotypes [126]. Moreover, the HVR sequences of the S1 and S2 genes have a specific site for positive selection pressure that allows for their receptor recognition or membrane fusion [127,128]. The positive selection occurred mainly once the viruses moved from the infected hosts to different hosts for adaptation or variation [129,130]. Clade-specific positive selection can permit the analysis of the different selection pressures on different clades in the phylogenetic tree [131]. Selective pressure can proceed to evolution or drive natural selection since some genes are unable to express themselves through replication; therefore, they cannot be transferred to the next generation. Some reasons for viral selection pressure arise from a series of immunization processes and incompletely protected birds. Likewise, the alteration in tissue tropism has been changed in the coding sequences of some CoVs [132].

The high occurrence of recombination has been noticed in the non-segmented RNA of CoVs and has occurred by several mechanisms, such as recombination by copy choice, replicative recombination without detachment, non-replicative recombination, re-assortment, and non-homologous recombination [133,134]. It has been reported that the genomes DE072 and D1466 may serve as recombination ‘hot spots’ and template switching sites for the viral encoded polymerase: they comprise the Gene 3, 4, 5, 6, and 3′ non-coding regions linked phylogenetically with diverse topology and specify the intergenic consensus sequence (CTGAAACA or CTTCACCA), which is a highly conserved sequence in this region. Recombination can support evolutionary benefits over generations to produce a diverse virus population that could be resistant to vaccination and the host’s immune system [135,136]. Some studies have made statements regarding the ability of a polymerase-
switched template to generate chimeric progeny during transcription [137,138]. Previous research has suggested that swine testis cells co-infected with TGEV and embryonated chicken eggs co-infected with different IBV strains are able to form multiple recombinant viruses [139,140]. The evolution of viruses through the acquisition of host-cell genes has been demonstrated in positive and negative sense viruses by homologous and non-homologous recombination. It has been reported that the acquisition of novel cellular genes and gene fragments of CoVs using overlapping reading frames for N and 9b genes has occurred through horizontal gene transfer [141,142].

4. PTM and Structural Motif of IBV

PTMs are the enzymatic modifications of proteins that are commonly synthesized by ribosomes translating mRNA into polypeptide chains. PTMs are various sets of transformations that assist in expanding the limited genomes of organisms. PTM plays a central role in regulating the folding of proteins; it also plays a part in the cellular functions by covalently adding to the specific functional, chemical, or complex group or by signaling activities of proteins in intricate signal transduction pathways [79]. Cleavage, palmitoylation, N-glycosylation, phosphorylation, and structural motifs, such as the leucine-rich repeat (LRR), are among the most common PTMs which are an important part of the IBV ecology, as shown in Figure 2. Therefore, it is essential to identify these modifications and structural motifs through bioinformatics.

![Post translation modification diagram](image)

Figure 2. Post-translational modification and structural motifs in various types of IBV protein. (E: Envelope protein, N: Nucleocapsid protein, M: Membrane protein, S: Spike protein, HE: hemagglutininesterase glycoprotein, Nsp: Nonstructural proteins).

The spike glycoprotein is typically translated as a precursor protein that includes a protease cleavage site motif to facilitate cleavage into S1 and S2 subunits, which are sensitive to furin or furin-like proteases [143–145]. Another notable trait that distinguishes COVID-19 from other animal CoV is the presence of a polybasic cleavage site, which facilitates the virus’s cleavage by furin and other proteases and influences host tropism [146]. The cleavage recognition site in the spike gene contains five basic amino acids that are cleaved by host-cell serine proteases [55,147,148]. Host-cell serine proteases catalyze hydrolysis reactions and this is the main reason that peptide bond cleavage occurs [149,150]. The initial identification of the cleavage site at the fourth and fifth basic amino acids suggests that a protease enzyme specifically recognizes and cleaves the protein at these positions, resulting in the formation of smaller protein fragments. The main purpose of such a site is to affect how much substrate fits the amino acid sequence around the bond to be cleaved [147]. Trypsin and chymotrypsin are two types of proteases clustered in the serine
protease family, where trypsin cleaves after lysine and arginine and chymotrypsin acts on tryptophan, tyrosine, and phenylalanine [151,152]. Numerous studies have suggested that CoV proteins undergo several types of PTMs that can significantly affect the viral replication and pathogenesis of the S protein [79,153]. The membrane topology of the CoV S protein was characterized by the TM protein (Figure 3A), which includes a large N-terminal ectodomain, a single TM domain, and a small C-terminal endodomain [154,155]. The luminal (virion exterior) ectodomain of CoV S protein is modified with N-linked glycosylation and disulfide bond formation, whereas conserved cysteine residues in the cytosolic tail are modified by palmitoylation (Figure 3B). Additionally, the N-linked glycosylation of complete S proteins can be cleaved, exposing the formation of trimers/dimers that are more immunogenic. The glycosylated S, M, E, and multiform 3a are viral membrane proteins, while the phosphorylated N protein is attached to viral genomic RNA in a beads-on-a-string style. Furthermore, cleavage of polyproteins of type 1a/1ab and PTM of different types of structural and nonstructural proteins give rise to mature and functional viral proteins [79].

![Diagram of CoV S protein](image)

**Figure 3.** The diagram represents the membrane topology and PTMs of the S protein from the three CoVs SARS-CoV, MHV, and IBV. (A) Membrane topology of CoV S protein. The crystal structures of the CoV protein covering the six-helix bundle fragments of S2 and the globular S1 domains are demonstrated. (B) Chief functional domains and PTMs on CoV S protein with groups of protein interactions comprising several types of modified residues are also specified. Endo: Endodomain, FP: Fusion peptide, HR: Heptad repeat, N-gly: N-glycosylation, Palm: Palmitoylation, RBD: Receptor-binding domain, S: Spike, TM: Transmembrane domain (adapted from Fung et al.) [79] (permission obtained).

The major neutralizing antigen in all CoV, except in monoclonal antibodies against the M protein, is the S glycoprotein. However, some epitopes or antigenic sites on the post-fusion S protein could induce non-neutralizing and precarious antibodies that enhance reinfecitivity [156,157]. For instance, Lontok et al. [158] reported that the S protein of IBV, a Group-3 CoV, is not transported to the cell surface but rather retained at an intracellular
portion; a similar statement has been documented for other CoVs of Groups 1 and 2. The intracellular transport of the IBV spike (S) protein is critical during the budding process at the ER–Golgi intermediate compartment (ERGIC) \[159,160\]. Therefore, the main structural proteins of IBV (S, M, E, and N located at the ERGIC membranes) allow for organized budding that is predisposed to allow virus assignment to target host cells \[161,162\]. Moreover, sialic acid plays a critical role as a receptor element for IBV infection by attaching receptors on erythrocytes, thus facilitating virus–receptor complex formation \[160\]. Similarly, IBV needs a relatively large amount of sialic acid on the cell surface to induce infection. Conversely, intracellular retention has also been observed for not only the IBV-S protein but also other CoVs, since the dityrosine motif in the cytoplasmic tail is responsible for the intracellular retention of surface proteins \[163\]. Moreover, tyrosine-containing tetrapeptides located in the cytoplasmic tails of many CoV S proteins, including IBV, are a critical component of the retention signal. The group of S-tail mutants includes those with the deletion of a dityrosine motif or point mutations in the S protein, where either of the tyrosines, or both, are substituted by alanine \[143\].

Palmitoylation is a post-translational process that involves the covalent attachment of saturated fatty acids, such as palmitic acid, to protein molecules at specific cysteine, serine, or threonine residues \[164,165\]. The functional effects of palmitoylation depend on the specific proteins involved; it plays important roles in subcellular trafficking, protein–protein interactions, and other biological processes \[166,167\]. The main site of palmitoylation is the C-terminal cysteine residues of membrane glycoproteins \[127\]. In CoVs, the S protein was first identified as palmitoylated in cells infected with MHV-A59 where it was found to contain palmitate (3H) \[168\]. This modification is typically found on the unglycosylated S protein in MHV-infected cells treated with tunicamycin \[169\]. Palmitoylation of S protein is associated with a reduced level of S protein complexing with virion membrane (M) proteins when S protein is removed from the virions \[170,171\]. However, palmitoylated S proteins can still be expressed on the cell surface during cell–cell fusion \[79\]. Some mutant CoV strains exhibit mutations in putative palmitoylation sites, resulting in reduced infectivity, and this supports the role of palmitoylation in virion assembly, fusion with cellular membranes, and the infection of cells \[168,172\]. In addition to the S protein, palmitoylation has also been observed on the CoV E proteins in IBV, SARS-CoV, and MHV as well as on other integral membrane proteins, where cysteine residues near the TM domain serve as targets for this modification \[173–175\]. Another study on the MHV S protein found that treatment with the palmitoyl acyltransferase inhibitor 2-bromopalmitate decreased the CoV infectivity \[176\].

Glycosylation of S and M proteins of CoVs contributes directly to fusion, receptor binding, activation, and antigenic features \[177,178\]. The M glycoprotein of the MHV strain M (MHV-M) is individually surrounded by O-linked oligosaccharides and confined to the Golgi region when expressed alone \[179\]. In fact, the viral proteins have O-linked glycans which are used in O-linked glycosylation and show that these peptides play a significant part in the biological function of CoV \[180\]. Even though MHV-M has been shown to have in vitro neutralizing activity and can protect mice from lethal virus challenges, it is still unclear whether anti-M antibodies have neutralizing activity against IBV-CoV and SARS-CoV \[181\]. Another study discovered that the M protein continued to have normal glycosylation following its administration even if tunicamycin inhibited the N-linked glycosylation of the S protein \[182\]. Glycosylation is essential for secondary protein processing inside the cell and changes in virulence and cellular tropism are associated with S and M proteins that are modified by glycosylation \[183\]. The first identification of N- and O-linked glycosylation in CoV S and M proteins was observed in the MHV (mouse hepatitis virus) receptor \[184\]. Similar findings were reported by Cavanagh et al. \[185\] and Reis et al. \[186\], indicating that the S glycoproteins of IBV are modified by N-linked glycosylation, which can impact the interaction between the virus and its receptors. Furthermore, N-linked glycosylation was also found in the endoplasmic reticulum (ER) resident of the MHV S-glycoprotein, leading to the synthesis of high mannose oligosaccharides \[187\].
process is influenced by monensin, a compound that affects glycosylation. Monensin is a Golgi transport blocker that interferes with the transport of the MHV-S protein from the trans-Golgi network to the cell surface. Some studies have suggested that the M protein of CoVs is N-linked glycosylated, containing a single glycosylation site, which can affect the virus’s interface with receptors [186,188]. The small luminal N terminus of IBV-E covers a consensus site for N-linked glycosylation, while SARS-CoV is expected to have two potential glycosylation sites [70,79].

Protein phosphorylation plays a vital role in regulating protein function and communicating signals inside the cell, mainly functioning on serine, threonine, or tyrosine amino acid residues [189]. According to a study by Emmott et al. [190], it was observed that the IBV-N protein undergoes phosphorylation throughout the early stages of infection. The level of phosphorylation detected in the N protein during this period was found to be 3.5-fold and 10-fold higher compared to the N protein obtained from infected tissue culture. Two phosphorylated clusters were recognized in IBV by mass spectroscopic methods, such as Ser190 and Ser192, along with Thr378 and Ser379. All clusters were situated in the C-terminal regions of the N protein that is linked with the diversity of viral RNA from non-viral RNA [191]. Wang et al. [192] stated that the level of phosphorylated protein kinase R is significantly suppressed in IBV-infected cells. The Nsp2 protein’s role as a weak antagonist is believed to interfere with the normal signaling and function of the protein kinase R antagonist, which is crucial for inhibiting viral replication during protein overexpression. Overall, protein phosphorylation is a dynamic and essential process for regulating protein function, enabling the precise control of cellular processes and responses to environmental stimuli. It is a fundamental mechanism employed by cells to modulate protein activity, signal transduction, and cellular behavior.

An LRR is a protein structural motif that consists of repeating 20–30 amino acid residues irregularly rich in the hydrophobic amino acid leucine. Reed et al. [193] reported that the leucine-rich nuclear export signal exists in the C-terminal region of the nucleocapsid protein (N) of IBV and interacts with the NALP3-LRR domain. It was evaluated with alanine replacement and deletion mutagenesis to explore the relative contributions of the leucine residues that can export nuclear protein. The LRR domain is confined to microbial proteins and is correlated with innate immunity [194]. LRR-containing proteins are also closely linked with numerous cellular mechanisms, including apoptosis, ubiquitin-related progressions, and nuclear mRNA passage [195,196]. Due to its versatility and ability to facilitate diverse interactions, the LRP is found in a wide range of proteins across various species, including cell surface receptors, transcription factors, and components of the immune system; its presence and structural arrangement enable proteins to participate in critical biological processes and play essential roles in cellular function and signaling pathways.

Ubiquitination and its complement have regulated the level and activity of cellular proteins during PTMs [197]. Three distinct enzymes, ubiquitin-activating enzymes, ubiquitin-conjugating enzymes, and ubiquitin ligases, are sequentially required for protein ubiquitination [182]. According to Yu et al. [198], the PLP of avian IBV possesses deubiquitinating activity and inhibits the host’s antiviral signaling cascade by breaking down proteins associated with polyubiquitin. All CoVs’ pathogenicity largely depends on the ubiquitin-proteasome system (UPS) [199]. For the most part, CoVs control the host cell UPS and UPS inhibitors are known to inhibit viral replication [200]. SARS-CoV [201] and IBV [202] encode only one functional PLP that processes virus-encoded large replicase polyproteins and deubiquitinating enzymes to cleave the isopeptide bonds found in polyubiquitin chains and interferes with IFN antagonism activities that inhibit the host antiviral reaction. Based on structural and enzymatic studies, Gamma CoV of IBV can function in proteolytic processing mediated by the PLP catalytic activity that is related to PLP2 encoded by other CoVs [203]. Similarly, the PLP of IBV can catalyze the proteolysis of Gly–Gly dipeptide bonds to release mature cleavage products. Overall, ubiquitin modifications play
5. Conclusions

IBVs remain a great challenge globally due to their economic consequences in the poultry industry. Key factors, such as genetic mutation and recombination, often contribute to the continuous emergence of many local and classical variants. Consequently, knowledge of the genomic structure, along with PTMs for IBV, is very significant in understanding the biological activity of protein biosynthesis and molecular mechanisms, as well as increasing the functional diversity of proteomics and identifying potential vaccine targets. Currently, many studies recognize the various types of PTMs of IBV that will help predict the pathogenesis of other related CoVs through molecular mechanisms. Despite the identified PTMs, a better understanding of the IBV biological mechanisms and the interactors for a particular infection may be required. PTM sites are functionally important because most protein-based drugs deliver therapeutic properties through diverse forms of PTM. Regarding the study of the translational application, more strategies are needed to extend our molecular knowledge of the PTMs of CoV proteins and the interference of the PTMs of host proteins which might enable the therapeutic intervention of the protein–protein interactions.

Author Contributions: Conceptualization, M.S.A.B., S.S. (Shafiquzzaman Siddiquee) and Z.A.; investigation, M.S.A.B., S.S. (Suryani Saallah) and S.S. (Subir Sarker); resources, S.S. (Shafiquzzaman Siddiquee) and Z.A.; writing—original draft preparation, M.S.A.B.; writing—review and editing, S.S. (Shafiquzzaman Siddiquee), K.F.R. and S.S. (Subir Sarker); Supervision, S.S. (Shafiquzzaman Siddiquee), Z.A. and S.M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by UMSGreat Research Grant Scheme (GUG108-1/2017) in University Malaysia Sabah (UMS) Malaysia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The study did not report any data.

Acknowledgments: The author would like to thank all of the scientists who have been directly or indirectly involving the IBV study in the world. The authors also like to acknowledge the University Malaysia Sabah (UMS) Malaysia for funding the Research Grant (UMSgreat) Scheme (GUG108-1/2017). Sarker is the recipient of an Australian Research Council Discovery Early Career Researcher Award (grant number DE200100367) funded by the Australian government.

Conflicts of Interest: The authors declare no conflict of interest.

References


98. Narayanay, K.; Ramirez, S.I.; Lokugamage, K.G.; Makino, S. Coronavirus nonstructural protein 1: Common and distinct functions in the regulation of host and viral gene expression. Virus Res. 2015, 202, 89–100. [CrossRef]


126. (Genetic Diversity and Phylogenetics of Avian Coronavirus in Egyptian Wild Birds. *Viruses* 2019, 11, 57. [CrossRef]


133. Domingo, E. Molecular Basis of Genetic Variation of Viruses: Error-Prone Replication. *Virus Genes* 2015, 46, 80. [CrossRef] [PubMed]


135. Xu, Y. Genetic diversity and potential recombination between ferret coronaviruses from European and American lineages. *J. Infect.* 2020, 80, 350–371. [CrossRef]


138. Lowry, K.; Woodman, A.; Cook, J.; Evans, D.J. Evans. Recombination in enteroviruses is a biphasic replicative process involving the generation of greater-than genome length ‘imprecise’ intermediates. *PloS Pathog.* 2014, 10, e1004191. [CrossRef]


141. Makarenkov, V.; Mazoure, B.; Rabusseau, G.; Legendre, P. Horizontal gene transfer and recombination analysis of SARS-CoV-2 genes helps discover its close relatives and shed light on its origin. *BMC Ecol. Evol.* 2021, 21, 5. [CrossRef]


152. Herman, R.A.; Song, P.; Mrisky, H.P. Trypsin cleavage sites are highly unlikely to occur in celiac-causing restricted epitopes. GM Crops Food 2019, 11, 67–69. [CrossRef]


170. Thorp, E.B.; Boscaino, J.A.; Logan, H.L.; Goletz, J.T.; Gallagher, T.M. Palmitoylations on murine coronavirus spike proteins are essential for virion assembly and infectivity. J. Virol. 2006, 80, 1280–1289. [CrossRef]

171. Li, Q.; Liu, Y.; Zhang, L. Cytoplasmic tail determines the membrane trafficking and localization of SARS-CoV-2 spike protein. Front. Mol. Biol. 2022, 9, 1004036. [CrossRef]


175. He, M.; Jenkins, P.; Bennett, V. Cysteine 70 of ankyrin-G is S-palmitoylated and is required for function of ankyrin-G in membrane domain assembly. J. Biol. Chem. 2012, 287, 43995–44005. [CrossRef]


187. Pande, S.; Rahardjo, A.; Livingston, B.; Mujacic, M. Monensin, a small molecule ionophore, can be used to increase high mannose levels on monoclonal antibodies generated by Chinese hamster ovary production cell-lines. Biotechnol. Bioeng. 2008, 100, 1383–1394. [CrossRef]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.