COVID-19 and iron dysregulation: distant sequence similarity between hepcidin and the novel coronavirus spike glycoprotein

Sepehr Ehsani 1,2

Abstract

The spike glycoprotein of the SARS-CoV-2 virus, which causes COVID-19, has attracted attention for its vaccine potential and binding capacity to host cell surface receptors. Much of this research focus has centered on the ectodomain of the spike protein. The ectodomain is anchored to a transmembrane region, followed by a cytoplasmic tail. Here we report a distant sequence similarity between the cysteine-rich cytoplasmic tail of the coronavirus spike protein and the hepcidin protein that is found in humans and other vertebrates. Hepcidin is thought to be the key regulator of iron metabolism in humans through its inhibition of the iron-exporting protein ferroportin. An implication of this preliminary observation is to suggest a potential route of investigation in the coronavirus research field making use of an already-established literature on the interplay of local and systemic iron regulation, cytokine-mediated inflammatory processes, respiratory infections and the hepcidin protein. The question of possible homology and an evolutionary connection between the viral spike protein and hepcidin is not assessed in this report, but some scenarios for its study are discussed.

Keywords: Coronavirus, COVID-19, Cysteine, Ferroportin, Hepcidin, IL-6, Inflammation, Iron, Pufferfish, Sequence similarity, Spike protein

Findings

Background

As of the beginning of October 2020, 189 countries and regions are tackling the challenge of the pandemic caused by the novel coronavirus, with more than 35 million confirmed cases of infection worldwide [23, 33]. Coronaviruses, first described in the 1960s [4, 82], are mostly present in birds and mammals, and there have thus far been seven known coronavirus infectious disease outbreaks in humans causing respiratory illness [62, 113]. The first two strains causing mild or common-cold-like symptoms are called 229E, NL63, OC43 and HKU1. The first two strains are in the ‘alpha’ coronavirus subgroup, whereas the latter two are ‘beta’ coronaviruses. The severe acute respiratory syndrome coronavirus (SARS-CoV-1) of 2002, the Middle East respiratory syndrome coronavirus (MERS-CoV) of 2012, and now the SARS-CoV-2 of 2019 (causing the ‘COVID-19’ disease that was declared a pandemic on 11 March 2020) are the remaining three known coronaviruses (all of the beta subgroup) causing severe human disease [6]. This positive-sense single-stranded RNA virus family possesses the structural proteins spike (S), membrane (M) and envelope (E) proteins, along with the nucleocapsid (N) protein. It also has the largest genome among RNA viruses [75].

Much research interest has been devoted to the spike (glyco) protein (forming the characteristic ‘corona’) and its importance in the development of vaccines and antivirals [17, 35]. The angiotensin-converting enzyme 2 (ACE2) on the surface of many different cell types in humans can act as a receptor for this virus.

© The Author(s). 2020 Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.
(ACE2) is thought to be its (main, but perhaps not exclusive) receptor on human host cells [132, 136]. The spike protein is formed of a receptor-binding subunit (S1), a membrane-fusion subunit (S2), a single-pass transmembrane (TM) domain, and a cytoplasmic/intracellular tail (CT) [75, 115]. Of note, the S1 domain has a similar fold as human galectins (galactose-binding lectins) [97]. Briefly, in terms of the putative primary function of the spike protein, Li comments that “because coronavirus must enter cells for replication, membrane fusion is the central function of coronavirus spikes” [75].

**Search for sequence similarity**

A basic question that might arise is: what exactly makes the pathobiology and disease course of these particular viruses unique? And, could it be that, in addition to viral replication inside the particular types of human host cells, other intracellular processes specific to these viruses are involved? (For a thematically related inquiry, see e.g., ref. [58].) Having this in mind, we wondered if there might be any sequence similarity (and thereby potential structural similarity) between the SARS-CoV-2 spike protein (which has 1273 amino acids [133, 134]) and any vertebrate protein(s).

A simple BLAST search does not reveal any similarities with human proteins. However, based on previous experience with the pufferfish Takifugu rubripes proteome [41, 42, 111] and its unique evolutionary history, we initially restricted the search to this species. Briefly, teleost ancestor species underwent an ancient whole-genome duplication event [125], and teleost species with well-annotated genomes such as the pufferfish can provide invaluable insights into the sequence evolution of genes with a clear phylogenetic linkage to mammalian genes (for a related commentary, see ref. [76]). Interestingly, a query using the full-length SARS-CoV-2 spike protein (accession no. YP_009724390.1) revealed a sole hit with the pufferfish hepcidin protein (XP_003965681.1; score: 32.7, E-value: 0.54). Given that SARS-like coronaviruses can be found in bats [133, 134], we also used a full-length bat coronavirus sequence (ANA96027.1) as the query, which showed a closer match with pufferfish hepcidin (score: 38.5, E-value: 0.005). Conversely, a BLAST search in the Coronaviridae family of viruses using the pufferfish hepcidin (XP_003965681.1) revealed the bat coronavirus spike protein (ANA96027.1) as the top hit (score: 38.5, E-value: 5e-04). The scores and E-values here are not meant to indicate any claims of statistical significance but are rather provided for the purpose of comparison.

This similarity between the spike protein and hepcidin is at the cytoplasmic tail [80] of the spike protein, or, depending on the exact domain delineation, perhaps at the junction between the TM and CT domains. A multiple sequence alignment of this sequence region (generated using the AlignX feature of Vector NTI Advance 11.0, Invitrogen, Carlsbad, CA, USA), using three coronavirus spike proteins and four hepcidin proteins (from pufferfish, bat and human) is illustrated in Fig. 1a.

The alignment depicts a number of conserved motifs, particularly between the first pufferfish hepcidin sequence (various pufferfish species have at least two hepcidin-like genes [55]) and the coronavirus spike proteins. In a sense, the pufferfish sequence seems to act as a ‘bridge’ between the coronavirus motifs and those in the human hepcidin sequence. This may not be surprising particularly given the evolutionary context of teleost species described earlier. The similar cysteine-rich motif takes the following form: ‘LXXXTXCCXCKGXXXCGXCC(R/K)F’. Of note, the eight cysteines of the mature human hepcidin in the similarity motif, and the aligned cysteines of the SARS-CoV-2 spike protein, are not all specifically coded by one of the two cysteine-coding codons (TGC and TGT). Both codons are present in the respective gene segments. Also, for comparison purposes, as the coronavirus envelope protein [112] contains a related ‘LCAYCCN’ motif [135], this sequence was added as the last line of the alignment.

Although this is a ‘distant’ and limited sequence similarity, it cannot be attributed to ‘chance’. The search that found hepcidin did not reveal a range of similarities with other teleost proteins. Moreover, there are many cysteine-rich protein sequences in teleosts and vertebrates in general, yet this similarity to the hepcidin gene family (not merely a one-off sequence) was unique and specific. How or why this similarity arose is then a separate, and potential follow-up, question. On this point, rather than framing the question as one having to do with a chance/random occurrence, it could more aptly be framed under concepts related to convergent/adaptive evolution versus common ancestry. Clearly, however, no conclusive claims about homology and sequence conservation can be made at present without a concerted investigation focused on this topic. Nevertheless, the similarity reported here raises a potential and intriguing question of whether there could be mimicry of human hepcidin (structural, functional or otherwise), perhaps inside the host cell, by the TM-CT junction of the spike protein. These possibilities should not be dismissed based on the mere phylogenetic gap between coronavirus and vertebrate species. Furthermore, the phylogenetic gap, and the absence of clear evolutionary homologies between genes that may otherwise have unexplored functional linkages, should not dissuade one from pursuing such connections. As a case in point, gene network and protein structure/function linkages between certain yeast (unicellular eukaryotes) and mammalian genes proved to be quite beneficial in...
investigating neurotoxicity in human cells [123]. The question now is if the linkage between the spike protein and hepcidin could be expanded, and what reasonable scenarios for its experimental validation could be envisaged.

The Heptacidin protein

Hepcidin is a small peptide hormone that was discovered in 2000/2001 [7, 63, 69, 96, 102], and initially named LEAP-1 (liver-expressed antimicrobial peptide). It has an antiparallel beta-sheet fold and contains four disulfide bonds, and is involved in iron trafficking and the host's response to infection [34]. In fact, it has been remarked by a number of commentators that "hepcidin is to iron, what insulin is to glucose" [57]. Clear hepcidin orthologs appear to be missing in birds and invertebrates [105]. The human hepcidin (encoded by the HAMP gene on chromosome 19q13) is an 84-amino-acid prepropeptide, leading to a mature 25-amino-acid peptide that is detectable in blood and urine [65]. The proprotein convertase furin has been demonstrated to cleave prohepcidin at a polybasic site [73, 124]. Of note, a furin-like cleavage site ('RRAR') has been reported to exist in the ectodomain of the SARS-CoV-2 spike protein, which is
absent in coronaviruses of the same clade [5, 28]. The protein-coding part of the HAMP gene is split over three exons, with the 25-amino-acid mature peptide occurring on the last exon. In terms of its putative function, Pren- tice notes that “although the hepcidin molecule does it- self possess some antimicrobial activity, this is rather weak compared to peptides such as defensins, and its primary contribution to innate immunity is via regula- tion of iron” [105]. In fish species, one of the two hepc- din paralogs has been shown to potentially possess antimi- crobial effects in innate immunity [78, 79]. Finally, hepcidin binds to and mediates the degradation of ferro- portin (encoded by the SLC40A1 / FPNT gene), the only known cellular iron exporter. The structural details of this interaction are being mapped and studied in ever more detail [18, 88, 108].

Hepcidin and the CoV spike protein

Available structures and structural predictions

There are a number of solved structures of hepcidin [65, 72]. An NMR structure of human hepcidin is depicted in Fig. 1b (visualized using 3-D Molecule Viewer, Vector NTI Advance 11.0, Invitrogen) including the locations of the four putative disulfide bonds. Of note, in addition to various hepcidin orthologs containing eight cysteines, four-cysteine variants have been described in notothe- nioid fish species [137]. Available solved structures of the coronavirus spike glycoprotein, as far as our search could reveal, mostly utilize expression constructs that stop just short of the TM domain. As noted, this is partly because the protein’s ectodomain is the main focus of studies on viral binding to host surface recep- tors [129, 131, 132]. For example, Wrapp, Wang and colleagues have reported the cryo-electron microscopy structure of ectodomain residues 1–1208 of the spike protein (trimers in the prefusion conformation) [132], but this excludes the TM and CT domains. It goes without saying that the inclusion of transmembrane domains would require complicated structural elucidation proto- cols, and even then, one may still not be able to solve the structure of the protein in its entirety.

Moreover, we would like to note that using the Pfam- A (ver. 32) structural/domain database [46] in the HHpred remote homology and structure prediction tool- kit [145], the coronavirus spike protein regions analyzed here show some predicted structural similarity to lipolysis-stimulated receptor (LSR) lipoprotein receptor family (PF05624) [139], and hepcidin sequences show some predicted structural similarity to the Sar8.2 protein family found in Solanaceae plants (PF03058) [2]. What, if any, significance these findings may hold is unclear at present. Of more importance right now would be the theoretical and/or actual elucidation of the structure of the spike protein TM-CT junction region and a comparison with the available hepcidin structures (the Pfam hepcidin entry, PF06446, currently references six PDB structures). Of note, a recent in silico study has reported on a predicted structural similarity and compati- bility between hepcidin and an allosteric site in the SARS-CoV-2 spike protein [32].

Post-translational modifications

Given the prominence of cysteines in the aligned motif (Fig. 1a), how are they utilized in the respective similar domains? At first pass, the usages appear to be different: as noted earlier, in hepcidin, the cysteines may give rise to a compact disulphide-bridged peptide [65, 89] (Fig. 1b), whereas in coronavirus spike glycoproteins, the cysteines in the TM-CT junction serve as palmitoylation acceptor residues [117] (Fig. 1c) that facilitate membrane fusion [24]. It should be mentioned, however, that ‘mini-hepci- din’ conjugated to palmitoyl groups have been synthet- sized and studied previously [106], but these do not occur naturally. As for the SARS-CoV, at least a portion of the palmitoylation of the spike protein has been re- ported to occur in a pre-medial Golgi compartment [81]. However, there is also the possibility of cross-disulphide bond formation with a non-homologous small cluster of cysteines within the envelope protein [135]. Moreover, if S-palmitoylation is a reversible and dy- namic process [118], it is to be determined if the spike protein junction cysteines might in fact have a different post-translational modification in the host cytoplasmic environment (although there is no evidence of this at the moment, and the reversibility of palmitoylation in viral spike proteins has not been reported [47]). That be- ing said, in the cited paper by McBride and Machamer, the authors conclude that the palmitoylation of the SARS-CoV spike protein “was not necessary for S pro- tein stability, trafficking or subcellular localization” nor “for efficient interaction with M protein” [81]. To what extent the post-translational modifications of the spike protein and hepcidin, be it furin cleavage, disulphide bonds or palmitoylation, are in any way similar in an intracellular context, remains to be examined.

Scenarios for the investigation of homology and evolutionary connections

As alluded to earlier, in terms of the possibility of an evolutionary connection between the spike protein and hepcidin, one could imagine a scenario whereby an an- cestral spike protein acquired a hepcidin-like sequence from a host organism, and the new sequence was palmi- toylated to aid with membrane association. Li points out that “the primordial form of coronavirus spikes might contain S2 only” [75], and the cytoplasmic motif features highlighted in the current report do not appear to be present in other class I viral membrane fusion proteins
(which include the influenza virus hemagglutinin [20, 66]), although we have not performed an exhaustive search. However, a number of questions might then arise under this scenario, such as the difference between the primary localizations of hepcidin (considering its putative interaction with extracellular and transmembrane regions of ferroportin [12, 18, 108]) versus the CT domain of the spike protein. Alternatively, it might be argued that perhaps a case of convergent sequence evolution is at play. For example, the influenza hemagglutinin glycoprotein appears to have a conserved ‘CXICX’ motif in its cytoplasmic tail domain [93, 128], and perhaps an ancestral spike protein with similar features convergently acquired hepcidin-like sequence motifs. These are of course speculations and remain open questions. Investigations pursuing these topics could also make use of studies that attempt to trace the evolutionary history of hepcidin itself [67, 137].

Potential leads for coronavirus research

Hepcidin-like motifs among different spike proteins

One of the first questions that may arise from this potential sequence similarity is whether the comparison is equal for all seven known human coronavirus strains, or if it is more limited to the severe disease-causing viruses. Figure 2 depicts a sequence alignment to start to answer this question. The alignment is restricted to the length of the mature human hepcidin protein, which is depicted at the bottom row. The first four spike protein sequences are comprised of the mild/asymptomatic strains, i.e., 229E and NL63 (alpha coronaviruses) and Hku1 and OC43 (beta coronaviruses). These are followed by the spike proteins of MERS-CoV, SARS-CoV-1 and SARS-CoV-2. The bat coronavirus sequence and the four hepcidins are depicted in the same order as in Fig. 1a. Paying particular attention to the region between the two conserved cysteines (marked by two black arrows), there appears to be less similarity in the motif between the first four spike proteins than the rest of the sequences. Specifically, for example, a ‘conserved’ glycine in the sixth position from the first overall-conserved cysteine (indicated with a red arrow) appears to be an important residue that groups together the three disease-causing spike proteins with the bat sequence and the two teleost hepcidin proteins. How, if at all, these differences play out in the cell might be an intriguing experimental direction.

What can also be noted from Fig. 2 is that although there are appreciable differences in the spike protein domains visualized in the figure between the MERS-CoV and the two SARS-CoV sequences, the SARS-CoV-1 and SARS-CoV-2 spike proteins are almost identical in this region (the tabulated amino acid identity and conservation values shown were calculated for the aligned region using the AlignX program). Therefore, based on the hypothetical link presented in this report, it might be reasonable to assume that any pathobiological differences between the two SARS strains would not be as a result of any differing biology attributable to the hepcidin similarity domain. For reference, the SARS-CoV-2 genome has been reported to be 96% identical to bat CoV (nucleotide identity), 80% to SARS-CoV-1, 55% to MERS-CoV and 50% to common cold CoV (e.g., the 229E and OC43 strains) [13]. Also of note in Fig. 2 is the percentage residue identity and conservation between Takifugu hepcidin and the corresponding bat coronavirus spike protein region (54% identity, 62% conservation), which are the same identity and conservation values between the two Takifugu hepcidin paralogs.

More broadly, the potential sequence link reported here points to the need to build on previous research on the cytoplasmic tail of the spike protein (e.g., refs. [80, 100, 101]) to better understand its distinct roles from the time of viral attachment to possible protein-protein interactions inside the host cell. Of the rare mutations currently reported in the SARS-CoV-2 spike protein (using patient-isolated genomic data in the GISAID repository), only one (P1263L) appears to be in the cytoplasmic tail domain [68], placing it outside of the similarity region reported here.

Moving back to the biology of the hepcidin protein itself, as noted previously, hepcidin binds to and mediates the degradation of the iron exporter ferroportin. If the sequence similarity reported here is actually playing a significant role at the cellular level, could it be that, although the cellular localizations appear to be different based on current knowledge, the SARS-CoV-2 spike protein cytoplasmic tail can partly mimic the structure of hepcidin and interact with ferroportin? Could the cytoplasmic tail even coordinate and bind iron? These remain to be investigated experimentally, but of note, Rishi and colleagues recently reported on the intracellular localization of ferroportin dimers, and concluded that both the carboxy- and amino-termini of the protein are intracellular [109]. As cited earlier, the details of hepcidin’s own interaction sites with ferroportin remain the subject of different structural determination projects [12, 18, 108]. Relatedly, and of interest, Neves and colleagues, using experiments on iron overload in European bass (Dicentrarchus labrax), have discussed the functional partnership between hepcidin and ferroportin from an evolutionary perspective and suggested that this may “open new possibilities for the pharmaceutical use of selected fish [...] hepcidins during infections, with no impact on iron homeostasis” [90, 91].

Notwithstanding the specificities of the hepcidin-ferroportin interaction, the sequence similarity reported here also points to a possible broader focus on iron
biology. Following calcium, oxygen and lead, iron has historically been one of the most studied elements in cell biology [40], and as such there is a vast body of (at times conflicting) literature to draw upon relating to iron in coronavirus infections. Other investigators have reviewed broad themes from iron biology in the context of COVID-19 [39, 78, 79], but here I will only touch upon salient features of iron biology that revolve around hepcidin. In the three subsections that follow, I will briefly discuss relevant and recent research with
perspectives on inflammation, hypoxia and diagnostics. Moreover, the main themes discussed in these segments have been summarized pictorially in Fig. 3.

**Hepcidin, iron biology and inflammation**

Foremost among the potential physiological connections related to the present discussion is the so-called ‘cytokine storm’ mediated by interleukin-6 (IL-6) reported in some COVID-19 patients [70, 84, 87, 138]. This is not, however, restricted to IL-6 and in fact elevated levels of a bundle of pro-inflammatory cytokines has been reported in severe COVID-19 cases [25, 26]. Indeed, researchers have proposed that “reduced innate antiviral defenses coupled with exuberant inflammatory cytokine production are the defining and driving features of COVID-19” [19]. In terms of the chronology of events, we could consider an initial “immune defense-based protective phase” and a subsequent “inflammation-driven damaging phase” in the disease [116].

How could these relate to hepcidin biology? First and foremost, an early set of findings on hepcidin was that it is regulated by anemia, hypoxia and inflammation [92]. A general understanding of this regulatory network is that inflammation brought about by infections increases hepcidin production, which in turn can lead to the anemia of inflammation [50, 52]. Hepcidin production in the liver is induced by IL-6 [51, 94], and it has been reported that hepatic heparan sulfate affects and regulates IL-6-stimulated hepcidin expression [103]. Furthermore, heparin, the anticoagulant glycosaminoglycan that is a highly sulfated form of heparan sulfate, has been shown to be a potent inhibitor of hepcidin expression [104]. Of interest, anticoagulant treatment has been reported to be effective in a subset of severe COVID-19 patients [120–122], and researchers have determined that the procoagulant transferrin [120–122] is upregulated in SARS-CoV-2 infections [83]. Related to such concerns with inflammation and coagulation in COVID-19, there have also recently been proposals centered on ACE2 and the vasopressor system protein bradykinin [53, 110]. These are various leads requiring further investigation.

It is pertinent to note that dysregulated hepcidin is a defining feature of hemochromatosis, a condition characterized by hepcidin deficiency, increased plasma iron levels and transferrin saturation [21]. In addition, a not-yet-fully-established link of relevance here is the observations of a Kawasaki-disease-like systemic vasculitis syndrome in children infected with the novel

---

**Fig. 3** Summary of salient facets of coronavirus spike protein and human hepcidin biology. On the top left of the figure, three trimeric spike proteins are depicted on a segment of the viral membrane, two in a figurative style and the middle trimer in a more detailed domain-specific format. One palmitate is shown attached to the cytoplasmic tail of the spike proteins, as per Fig. 1c. Following the binding step to ACE2, a stylized viral fusion step is depicted based on refs. [120–122], where an early fusion scenario (as opposed to endocytosis) is envisioned. The assembly of new viral spike proteins is shown to be taking place in an ER-Golgi intermediate compartment. On the right half of the figure, various elements of hepcidin and iron biology are depicted, which could result in potential outcomes such as coagulopathy, ferroptosis and hyperferritinemia. In terms of hepcidin’s binding to ferroportin, two possible scenarios are depicted, whereby hepcidin could bind to the extracellular face of ferroportin or deeper inside the transmembrane cavity. Lastly, the geometric shapes for the spike and ACE2 proteins were adapted from ref. [119], for ferroportin from ref. [30], and for ferritin from ref. [29], all with permission from the publisher (Springer Nature)
coronavirus [64, 127], which is also being called multi-system inflammatory syndrome in children (MIS-C) [37, 44]. Incidentally, an association between a novel human coronavirus and Kawasaki disease was reported in 2005 [43], although other investigators were apparently not successful in confirming the link [38]. Nevertheless, if the association with the Kawasaki-disease-like syndrome is real, then it is noteworthy that increased hepcidin levels have been suggested as a biomarker for Kawasaki disease [61].

**Hepcidin, iron biology and hypoxia**

As opposed to the inflammation-based upregulation of hepcidin, anemia and hypoxia are usually taken to have the opposite effect on hepcidin’s expression [92]. Further to cytokine storms and characteristic immune reactions, hypoxic respiratory failure could be another major warning sign in COVID-19 patients [31]. Less established, but still important to mention, is also the debate surrounding the issue of hypoxia/hypoxemia and certain symptoms resembling, but differentiable from, altitude illness [8, 9, 15, 54]. Congruously, hepcidin expression levels have also been investigated in the context of high-altitude acclimatization [59] (see also ref. [27]). Moreover, hepcidin levels are known to increase in patients with acute respiratory distress syndrome (ARDS) [48, 86].

It may also be important to point to a number of circumstantial but perhaps important findings in the literature pertaining to lung disease. These include (i) a link between SARS and liver function abnormalities [74], (ii) the association of pulmonary iron overload and restrictive lung disease [49, 90, 91], (iii) the role of iron in pulmonary fibrosis [3], (iv) hepcidin’s modulation of the proliferation of pulmonary artery smooth muscle cells [107], and (v) the vital role of hepcidin in alveolar macrophage function [98]. Furthermore, (vi) hepcidin upregulation along with serum iron reduction has been reported in influenza infections [10, 45]. Importantly, however, iron dysregulation changes may only be at a local cellular/tissue level and not reach a systemic response [71], and may demonstrate selective tissue tropisms during different viral infections [11].

**Hepcidin, iron biology and diagnostics**

Presently, dexamethasone and other corticosteroids are among the very few widely-used treatments for COVID-19 [130], and it would be of great significance to have robust diagnostic correlates of the course of the disease. Could systemic changes in serum iron levels [60, 114, 126, 141] (with a possible view on the degree of abnormally normal serum ferritin in patients [25, 26]) or levels of hepcidin itself be detected in patients with varying COVID-19 severities? First, on the latter point, researchers have recently reported that increased serum levels of hepcidin and ferritin are indeed associated with the severity of COVID-19 [143, 144]. This combined hepcidin/ferritin diagnostic might potentially be promising. Serum ferritin levels alone are usually considered a general indicator of inflammation and infection, and as pointed out by Baron and colleagues, “the use of ferritin to diagnose iron deficiency may be problematic in patients with COVID-19 disease, who may have normal or high ferritin levels despite very low iron stores” [14, 143, 144]. In fact, COVID-19 has been proposed to be a part of the hyperferritinemic spectrum of conditions [56, 99].

As for correlations with plasma iron levels, Shah and colleagues have reported that “compared with patients with non-severe hypoxemia, patients with severe hypoxemia had significantly lower levels of serum iron”, and that “the association of serum iron with lymphocyte counts could reflect the requirement of the adaptive immune response for iron and may contribute to possible T cell dysfunction reported in COVID-19” [114]. The authors did not find significant differences in transferrin saturation or serum ferritin levels between the non-severe and severe hypoxemia patient groups. Congruently, a retrospective study of COVID-19 patients and serum iron deficiency has found that “the severity and mortality of the disease was closely correlated with serum iron levels” [142].

Beyond the immediate issue of diagnostic markers, two broader questions that could follow from the present work can be raised: first, does the spike protein, similar to hepcidin, potentially promote iron sequestration in (alveolar) macrophages [85] and hence impede the host’s initial immunological response? Second, could reports of the common presence of digestive symptoms in COVID-19 patients (e.g., ref. [95]) be explainable in part by a link to hepcidin? With respect to the biology surrounding these gastrointestinal symptoms, it is important to point out that a recent study concluded that in gut epithelial cells, the “expression of two mucosa-specific serine proteases, TMPRSS2 and TMPRSS4, facilitated SARS-CoV-2 spike fusogenic activity and promoted virus entry into host cells” [140]. It is also known that the liver-specific serine protease TMPRSS6 (matriptase-2) negatively modulates hepcidin [16, 22, 36]. Again, what if any significance these connections might hold remains to be determined.

Overall, the observations in this report suggest that, as a starting point, serum iron status would be a critical data category to be systematically collected from patients at various stages of the disease’s progression. Furthermore, iron dyshomeostasis in the case of COVID-19 may be a more specific pathobiological feature of this infection, and we can only know the answer to this scenario if more data related to iron biology is collected.
Table 1  Points of similarity and divergence between hepcidin and the coronavirus spike protein cytoplasmic domain based on current knowledge of the viral protein and its pathobiology (see main text for references)

| Potential Similarities | Potential Differences |
|------------------------|-----------------------|
| 1. Protein sequence: A unique but restricted sequence similarity exists between mature hepcidin and the cysteine-rich cytoplasmic tail of coronavirus spike proteins. | 1. Protein length and domains: Coronavirus spike proteins are much larger, multi-domain, proteins compared to the relatively short-length hepcidin proteins. |
| 2. Post-translational processing: The proprotein convertase furin cleaves prohepcidin, and has been reported to also activate (the ectodomain of) the SARS-CoV-2 spike protein. | 2. Post-translational modification: Cysteines in the cytoplasmic tail of coronavirus spike proteins are thought to be palmitoylation acceptor residues, and this modification has thus far not been reported to be reversible. Cysteines in hepcidin are used to form disulfide bonds. |
| 3. Cytokine storm: IL-6-mediated inflammatory responses have been reported in COVID-19 patients. Hepcidin production in the liver is induced by IL-6 and is well-studied in the context of the anemia of inflammation. In addition, COVID-19 may be associated with a Kawasaki-disease-like systemic vasculitis manifestation in children, and hepcidin levels have also been suggested as a biomarker for Kawasaki disease. | 3. Localization: Hepcidin is thought to interact with its main (iron-bound) interactor, ferroportin, extracellularly, whereas the spike protein cytoplasmic tail does not face the environment outside the viral membrane. (However, the cytoplasmic tail associates with the plasma membrane itself aided by its palmitoylation modifications, and hepcidin may interact with ferroportin close to ferroportin’s transmembrane regions in addition to extracellular residues.) |
| 4. Hypoxia/Hypoxemia: COVID-19 may lead to symptoms resembling, but differentiable from, altitude illness, and hepcidin expression levels have also been studied in the context of high-altitude acclimatization. | |

during the current pandemic. Could, for instance, clinical evidence be gathered to clarify if disorders of iron homeostasis might exacerbate symptoms in COVID-19 patients? As a final note, given the many known and yet-to-be-discovered intricacies of iron homeostasis in the body, research on therapeutic strategies that, for example, propose to utilize iron chelation or hepcidin antagonists should proceed cautiously [1, 52].

Conclusion
Theoretical analyses in new areas may necessarily entail reasonable speculations based on limited or disparate data. This is expected, but one should remain cognizant of overinterpretation, and pursue a rational course of theoretical inquiry to hopefully inform subsequent experimental investigations. Here, a purposeful and restricted protein sequence search revealed a potential sequence similarity between the relatively less-studied cysteine-rich cytoplasmic domain of coronavirus spike proteins and the vertebrate hepcidin protein. This is quite unlikely to be a spurious and random similarity. There are many cysteine-rich protein sequences in vertebrates, but the motif identified here is unique and specific, and also appears to tentatively set apart the disease-causing strains from the milder coronavirus strains. Following from this link, a number of emerging clinical strands of evidence (summarized in Table 1) were discussed which further link a biology surrounding hepcidin with coronavirus-caused pathobiology. While each piece of clinical evidence discussed does not by itself provide overwhelming corroboration for the hypothesis of the paper, the totality of evidence presented we believe make a strong case that if sequence and/or structural mimicry to hepcidin is taking place upon viral attachment to and entry in the host cell, then perhaps a local disease condition resembling iron dysregulation (e.g., iron overload) might ensue in the infected tissue(s). This hypothesis can be immediately tested in one of three ways. First, in the clinic, levels of various serum markers for iron biology could be more systematically measured and analyzed. This strand of investigation has already produced corroborating evidence in the form of increased serum levels of hepcidin and significantly lower levels of serum iron in COVID-19 patients. Second, computational investigations could examine potential structural mimicry between the two proteins and explore the effect of differing post-translational modifications. And third, the potential link to hepcidin could be relied upon in cell-based assays to determine the possibility of the involvement of the spike protein in iron biology.

Acknowledgements
The first draft of this manuscript was posted on 27 March 2020 on the arXiv preprint server (arxiv.org/abs/2003.12191v1). Thanks are due to Gerold Schmitt-Ulms (University of Toronto) and Farnaz Fahimi Hanzaee (University College London) for very helpful discussions.

Author’s contributions
S.E. performed the analysis and wrote the paper. The author(s) read and approved the final manuscript.

Author’s information
S.E. holds a Ph.D. in laboratory medicine and pathobiology from the University of Toronto, and was a postdoctoral fellow at the Whitehead Institute for Biomedical Research and the Computer Science and Artificial Intelligence Laboratory at the Massachusetts Institute of Technology. He is currently a Ph.D. student at the Department of Philosophy at University College London.

Funding
The author is supported by a University College London Overseas Research Scholarship.

Availability of data and materials
All sequences utilized are from publicly available repositories. Accession numbers are provided in the relevant sections.
iron uptake and hereditary hemochromatosis. J Biol Chem. 2002;277(40): 37597–603. https://doi.org/10.1074/jbc.M205305200.

4. Jones VG, Mills M, Suarez D, Hogan CA, Yeh D, Segal JB, et al. COVID-19 and Kawasaki disease: novel virus and novel case. Hosp Pediatr. 2020;10(6):537– 40. https://doi.org/10.1542/hpeds.2020-0123.

5. Jordan JB, Poppe L, Haniu M, Anveden T, Syed R, et al. Hepcidin revisited, disulfide connectivity, dynamics, and structure. J Biol Chem. 2009; 284(36):24155–67. https://doi.org/10.1074/jbc.M109017764.

6. Kielian M. Mechanisms of virus membrane fusion proteins. Annu Rev Virol. 2014;1(1):17–89. https://doi.org/10.1146/annurev-virology-031414-085521.

7. Kim CH, Kim EJ, Nam YK. Chondrostean sturgeon hepcidin: an evolutionary link between tetecost and tetrapod hepcidins. Fish Shellfish Immunol. 2019; 88:117–25. https://doi.org/10.1016/j.fsi.2019.02.045.

8. Korber B, Fischer WM, Granakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell. 2020;182(4):121–27e19. https://doi.org/10.1016/j.cell.2020.06.043.

9. Krause, A., Netzer, V., Magert, H. J., Schulz, A., Frensmann, W. G., Schulz- Knappe, P., & Adamm, K. (2000). LEAP-1, a novel highly disulfide-bonded human protein, exhibits antimicrobial activity. FEMS Lett, 480(2–3), 147–150. https://doi.org/10.1016/S0168-6444(00)01920-7.

10. Laing AG, Lorenz A, Del Molino Del Barrio I, Das A, Fish M, Monni L, et al. A dynamic COVID-19 immune signature includes associations with poor prognosis. Nat Med. 2020. https://doi.org/10.1038/s41591-020-1038-6.

11. Lakhal-Littleton S, Crosby A, Frise MC, Mohammad G, Carr CA, Lock PAM, et al. Intracellular iron deficiency in pulmonary arterial smooth muscle cells induces pulmonary arterial hypertension in mice. Proc Natl Acad Sci U S A. 2019;116(26):13122–30. https://doi.org/10.1073/pnas.1820101116.

12. Lauth X, Babon JJ, Stannard JA, Singh S, Nizet V, Carlborg JM, et al. Bass hepcidin synthesis, solution structure, antimicrobial activities and synergism, and in vivo hepatic response to bacterial infections. J Biol Chem. 2005; 280(10):9272–82. https://doi.org/10.1074/jbc.M411511200.

13. Lee P. Commentary to “Post-translational processing of hepcidin in human hepatocytes is mediated by the prohormone convertase furin,” by Erika Valore and Tomas Ganz. Blood Cells Mol Dis. 2008;40(1):139–40. https://doi.org/10.1016/j.bcmd.2007.07.010.

14. Leffkowitch JH. Hepatobiliary pathology. Curr Opin Gastroenterol. 2005;21(3): 260–9. https://doi.org/10.1097/01.mog.0000195820.78532.68.

15. Li F. Structure, function, and evolution of coronavirus spike proteins. Annu Rev Virol. 2016;3(1):237–61. https://doi.org/10.1146/annurev-virology-110615-042301.

16. Lidgard S, Love AC. Rethinking living fossils. Bioscience. 2018;68(10):760–70. https://doi.org/10.1093/biosci/bby084.

17. Linder ME, Deschenes RI. Palmitoylation: policing protein stability and traffic. Nat Rev Mol Cell Biol. 2007;8(1):74–84. https://doi.org/10.1038/ nm2084.

18. Liu B, Liu GD, Guo HY, Zhu KC, Guo L, Zhang N, et al. Characterization and functional analysis of liver-expressed antimicrobial peptide-2 (LEAP-2) from golden pompano Trachinotus ovatus (Linnaeus 1758). Fish Shellfish Immunol. 2020;104:419–30. https://doi.org/10.1016/j.fsi.2020.06.029.

19. Liu W, Zhang S, Nekhai S, Liu S. Depriving iron supply to the virus represents a promising adjuvant therapeutic against Viral Survival. Curr Clin Microbiol Rep. 2020-1. https://doi.org/10.1007/s40279-020-00140-w.

20. Mcbride CE, Li J, Machamer CE. The cytokinetic tail of the severe acute respiratory syndrome coronavirus spike protein contains a novel endoplasmic reticulum retrieval signal that binds COPI and promotes interaction with membrane protein. J Virol. 2020;84(5):2418–28. https://doi.org/10.1128/jvi.02466-05.

21. Mcbride CE, Machamer CE. Palmitoylation of SARS-CoV-2 protein is necessary for pairing into detergent-resistant membranes and cell-cell fusion but not interaction with M protein. Virology. 2010;405(1):139–48. https://doi.org/10.1016/j.virol.2010.05.031.

22. Mcintosh K. Commentary: Mcintosh K, Chao RK, Krause HE, Wasi R, Moega HE, Mufson MA. Coronavirus infection in acute lower respiratory tract disease of infants. J Infect Dis 1974; 130:502–7. J Infect Dis. 2004;190(5):1033– 41. https://doi.org/10.1086/422961.

23. McClaughlin KM, Bechtel M, Bokjasova D, Munch C, Ciesek S, Wass MN, et al. COVID-19-related coagulopathy is transferrin missing link? Diagnostics (Basel). 2020;10(8). https://doi.org/10.3390/diagnostics10080539.

24. Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JC, et al. COVID-19: consider cytokine storm syndromes and immunosuppression.
98. Perez E, Baker JR, Di Giandomenico S, Kermani P, Parker J, Kim K, et al.
96. Park CH, Valore EV, Waring AJ, Ganz T. Hepcidin, a urinary antimicrobial
95. Pan L, Mu M, Yang P, Sun Y, Wang R, Yan J, et al. Clinical characteristics of
92. Nicolas G, Chauvet C, Viatte L, Danan JL, Bigard X, Devaux I, et al. The gene
90. Neves J, Leitz D, Kraut S, Brandenberger C, Agrawal R, Weissmann N, et al.
89. Nemeth E, Ganz T. The role of hepcidin in iron metabolism. Acta Haematol.
88. Muriuki JM, Mentzer AJ, Band G, Gilchrist JJ, Carstensen T, Lule SA, et al. The
85. Michels K, Nemeth E, Ganz T, Mehrad B. Hepcidin and host defense against
80. Muriuki JM, Mentzer AJ, Band G, Gilchrist JJ, Carstensen T, Lule SA, et al. The
28. http://jnnp.bmj.com/content/180/6/637.short
27. https://doi.org/10.1136/jnnp.2019.361514
26. https://www.nature.com/articles/s41396-019-03620-z
25. https://doi.org/10.1111/1365-2567.12926
24. https://doi.org/10.1016/j.ijmedchem.2015.06.006
23. https://doi.org/10.1016/j.virol.2007.03.012
22. https://doi.org/10.1111/j.1365-2958.2006.03134.x
21. https://doi.org/10.1186/1471-2121-9-19
20. https://doi.org/10.1186/1471-2156-9-43
19. https://doi.org/10.1016/S0022-2010(99)80010-7
18. https://doi.org/10.1016/j.virol.2006.04.021
17. https://doi.org/10.1016/j.ijmedchem.2015.09.005
16. https://doi.org/10.1016/j.ijmedchem.2015.06.007
15. https://doi.org/10.1007/s10021-015-0153-2
14. https://doi.org/10.1016/j.virol.2005.09.006
13. https://doi.org/10.1016/j.ijmedchem.2015.09.003
12. https://doi.org/10.1016/j.ijmedchem.2015.08.005
11. https://doi.org/10.1016/j.ijmedchem.2015.06.013
10. https://doi.org/10.1007/s10021-015-0152-3
9. https://doi.org/10.1016/j.ijmedchem.2015.09.002
8. https://doi.org/10.1016/j.ijmedchem.2015.08.004
7. https://doi.org/10.1016/j.ijmedchem.2015.06.012
6. https://doi.org/10.1016/j.ijmedchem.2015.06.011
5. https://doi.org/10.1016/j.ijmedchem.2015.06.010
4. https://doi.org/10.1016/j.ijmedchem.2015.06.009
3. https://doi.org/10.1016/j.ijmedchem.2015.06.008
2. https://doi.org/10.1016/j.ijmedchem.2015.06.007
1. https://doi.org/10.1016/j.ijmedchem.2015.06.006
126. van Swelm RPL, Wetzels JFM, Swinkels DW. The multifaceted role of iron in renal health and disease. Nat Rev Nephrol. 2020;16(2):77–98. https://doi.org/10.1038/s41581-019-0197-5.

127. Verdoni L, Mazza A, Gervasoni A, Martelli L, Ruggeri M, Cluffreda M, et al. An outbreak of severe Kawasaki-like disease at the Italian epicentre of the SARS-CoV-2 epidemic: an observational cohort study. Lancet. 2020;395(10239):1771–8. https://doi.org/10.1016/S0140-6736(20)31103-X.

128. Wagner R, Henwig A, Azzouz N, Klenk HD. Acylation-mediated membrane anchoring of avian influenza virus hemagglutinin is essential for fusion pore formation and virus infectivity. J Virol. 2005;79(10):6449–58. https://doi.org/10.1128/JVI.79.10.6449-6458.2005.

129. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell. 2020;181(2):281–92e286. https://doi.org/10.1016/j.cell.2020.02.058.

130. WHO Rapid Evidence Appraisal for COVID-19 Therapies Working Group, Sterne JAC, Murthy S, Diaz JV, Slutsky AS, Villar J, et al. Association between Administration of Systemic Corticosteroids and Mortality among Critically ill Patients with COVID-19: a meta-analysis. JAMA. 2020. https://doi.org/10.1001/jama.2020.17023.

131. Wrapp D, McLellan JS. The 3.1-angstrom Cryo-electron microscopy structure of the porcine epidemic diarrhea virus spike protein in the Prefusion conformation. J Virol. 2019;93(23). https://doi.org/10.1128/JVI.00923-19.

132. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020;367(6483):1260–3. https://doi.org/10.1126/science.abb2507.

133. Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, et al. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. Cell Host Microbe. 2020;27(3):325–8. https://doi.org/10.1016/j.chom.2020.02.001.

134. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579(7798):265–9. https://doi.org/10.1038/s41586-020-2008-3.

135. Xu Q, Cheng CH, Hu P, Ye H, Zou L, Cao L, et al. Adaptive evolution of hepcidin genes in antarctic nototheniid fishes. Mol Biol Evol. 2008;25(6):1099–112. https://doi.org/10.1093/molbev/msn056.

136. Xu Q, Chen Y, Ji Y, He X, Liu Y, et al. The E protein is a multifunctional membrane protein of SARS-CoV. Genomics Proteomics Bioinformatics. 2003;1(2):131–144. https://doi.org/10.1016/s1672-0404(03)10179-9.

137. Xu X, Han M, Li T, Sun W, Wang D, Fu B, et al. Effective treatment of severe COVID-19 patients with tocilizumab. Proc Natl Acad Sci U S A. 2020;117(20):10970–5. https://doi.org/10.1073/pnas.2005615117.

138. Xu X, Han M, Li T, Sun W, Wang D, Fu B, et al. Effective treatment of severe COVID-19 patients with tocilizumab. Proc Natl Acad Sci U S A. 2020;117(20):10970–5. https://doi.org/10.1073/pnas.2005615117.

139. Zahorik AM, Dioudidou K, Kim J, Deblasio R, Zhou J, et al. TMPRSS2 and TMPRSS4 promote SARS-CoV-2 infection of human small intestinal enterocytes. Sci Immunol. 2020;5(47). https://doi.org/10.1126/sciimmunol.abc3582.

140. Zhang L, Liu Y. Potential interventions for novel coronavirus in China: a systematic review. J Med Virol. 2020;92(5):979–90. https://doi.org/10.1002/jmv.25707.

141. Zhao K, Huang J, Dai D, Feng Y, Liu L, Nie S. Serum iron level as a potential predictor of coronavirus disease 2019 severity and mortality: a retrospective study. Open Forum Infect Dis. 2020;7(10):ofaa250. https://doi.org/10.1093/ofid/ofaa250.

142. Zhou C, Chen Y, Ji Y, He X, He X, et al. Increased serum levels of Hepcidin and ferritin are associated with severity of COVID-19. Med Sci Monit. 2020;26:e926178. https://doi.org/10.12659/MSM.926178.

143. Zimmermann L, Stephens A, Naim SZ, Rau D, Kubler J, Lozajic M, et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its Core. J Mol Biol. 2018;430(15):2237–43. https://doi.org/10.1016/j.jmb.2017.12.007.