Short linear motif candidates in the cell entry system used by SARS-CoV-2 and their potential therapeutic implications

Bálint Mészáros1*, Hugo Sámano-Sánchez2†, Jesús Alvarado-Valverde1,2, Jelena Čalyševa1,2, Elizabeth Martínez-Pérez1,3, Renato Alves1, Denis C. Shields4, Manjeet Kumar1*, Friedrich Rippmann5, Lucía B. Chemes6*, Toby J. Gibson1*

The first reported receptor for SARS-CoV-2 on host cells was the angiotensin-converting enzyme 2 (ACE2). However, the viral spike protein also has an RGD motif, suggesting that cell surface integrins may be co-receptors. We examined the sequences of ACE2 and integrins with the Eukaryotic Linear Motif (ELM) resource and identified candidate short linear motifs (SLiMs) in their short, unstructured, cytosolic tails with potential roles in endocytosis, membrane dynamics, autophagy, cytoskeleton, and cell signaling. These SLiM candidates are highly conserved in vertebrates and may interact with the µ2 subunit of the endocytosis-associated AP2 adaptor complex, as well as with various protein domains (namely, I-BAR, LC3, PDZ, PTB, and SH2) found in human signaling and regulatory proteins. Several motifs overlap in the tail sequences, suggesting that they may act as molecular switches, such as in response to tyrosine phosphorylation status. Candidate LC3-interacting region (LIR) motifs are present in the tails of integrin β3 and ACE2, suggesting that these proteins could directly recruit autophagy components. Our findings identify several molecular links and testable hypotheses that could uncover mechanisms of SARS-CoV-2 attachment, entry, and replication against which it may be possible to develop host-directed therapies that dampen viral infection and disease progression. Several of these SLiMs have now been validated to mediate the predicted peptide interactions.

INTRODUCTION

The coronavirus disease 19 (COVID-19) pandemic is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), an enveloped, single-stranded RNA virus. It had infected more than 68 million people and caused over 1.5 million deaths globally by mid-December 2020. SARS-CoV-2 belongs to the Coronavirus family, whose members are common human pathogens responsible for the common cold, as well as for some emerging severe respiratory diseases. Among them are the SARS-CoV and the Middle East respiratory syndrome coronavirus (MERS-CoV), the former of which caused over 8000 cases in 2003 with a fatality rate of ~10% and the latter caused about 2500 infections in 2012 with a fatality rate of 37% (1). Another coronavirus, infectious bronchitis virus (IBV), infects birds and has been used as a model in coronavirus research (2). SARS-CoV-2, like SARS-CoV (3), uses the angiotensin-converting enzyme 2 (ACE2) as a receptor (4–6) to attach to host cells. ACE2 is a single-pass type I membrane protein with a short cytosolic C-terminal region for which the functionality, however, is mostly unknown.

Earlier results show that the SARS-CoV-2 receptor-binding domain (RBD) of the spike protein interacts with ACE2 for cellular entry. In 2004, ACE2 was shown to be highly expressed in lungs by anti-ACE2 antibody staining (7). However, several 2020 papers using both antibodies and single-cell mRNA sequencing now find that there is very little ACE2 gene expression in normal lungs (8–11). This suggests that the ACE2 receptor is insufficient to establish severe lung disease and that SARS-CoV-2 can bend other cell surface receptors on human lung cells. One group of candidate co-receptors are the integrins that bind a large variety of ligands harboring an RGD (Arg-Gly-Asp) sequence motif, as recent analysis of the RBD identified a possibly functional RGD motif (12).

Integrins are major cell attachment receptors, which are known to be targeted by a range of viruses—including HIV, herpes simplex virus-2, Epstein-Barr virus (EBV), and the foot and mouth disease virus (FMDV)—for cell entry and activation of linked intracellular pathways (13–15). Integrins are special types of receptors, as they propagate signals in both directions; extracellular ligands can induce cytoplasmic pathway activation, but intracellular interactions with the cytosolic tails can influence the structure of the ectodomains and hence ligand-binding affinity. The complexity of integrin signaling stems from the dimeric structure of integrins, as they are composed of two subunits, α and β. For the RGD-binding integrins, the ligand-binding surface lies at the interface of the two integrin subunits, with both subunits making contacts with the ligand. These RGD motifs are recognized by at least 8 of the 24 human integrins, and the flanking residues next to the core RGD motif are known to play a decisive role in selectivity (16). Several viral proteins contain RGD (or RGD-like) short linear motifs (SLiMs) for integrin modulation; in addition, not only some viruses can use integrins on the host cell surface but also HIV/SIV (simian immunodeficiency virus) can incorporate integrins into their own membranes for mediating interactions with the host (17). Therefore, integrins can potentially be targeted...
at both the extracellular and the intracellular side to combat pathogenic hijacking.

Viruses, as obligate intracellular entities, need to interfere with major cellular processes like vesicular trafficking, cell cycle, cellular transport, protein degradation, or signal transduction to satisfy their replication, enzymatic, metabolic, and transport needs (18). To achieve this, a large number of host processes are hijacked using SLiMs often located in intrinsically disordered regions to establish protein–protein interactions with host proteins or undergo post-translational modifications (PTMs) such as tyrosine phosphorylation. For example, cellular signaling relies heavily on the use of SLiMs (19, 20). The low affinity and cooperativity of SLiM-based molecular processes allow reversible and transient interactions that can work as switches between distinct functional states and are regulated in both time and space (21, 22). Conditional switching of SLiMs, for example, through phosphorylation, can induce the exchange of binding partners for a protein, thus mediating molecular decision-making in response to signals reporting on the cell state (20). The Eukaryotic Linear Motif (ELM) resource (http://elm.eu.org/) is a dedicated database and exploratory server for over 280 manually curated SLiM classes with experimental evidence, each of them defined by a POSIX regular expression (23).

As explained above, a major strategy of viruses is to abuse the host system by using mimics of eukaryotic SLiMs to compete with extracellular or intracellular binding partners or to sequester host proteins (18). This dependence of viruses and many other pathogens on SLiM-mediated functions suggests that there is an opportunity to drug the cell systems where these interactions are being hijacked (24). For example, tyrosine kinase inhibitors, often used in anticancer therapy, have shown promising coronavirus replication inhibition in infectious cell culture systems (2, 25–27). In the remainder of the introduction, we will describe some of the major pathways hijacked by viruses to accomplish cell attachment, entry, and replication, which are suggested by our results to be relevant to SARS-CoV-2 infection.

Receptor-mediated endocytosis (RME) is a cellular import process triggered by cell surface receptor proteins, including any cargoes attached to them, in which a large vesicular structure is assembled entirely through cooperative low-affinity interactions of SLiMs and phospholipid head groups with their globular protein domain partners. The vesicles are strong and stable, yet flexible and dynamically assembled and disassembled. The external triggering of surface receptors (many of which have the YxxPhi or NPxY tyrosine sorting motifs) is transmitted across the plasma membrane, inducing local enzymatic modification of lipid head groups from phosphatidylinositol-4-phosphate (PI4P) to phosphatidylinositol 4,5-bisphosphate (PI(4,5)P2) by the PIPK1 kinase. The local enrichment of PI(4,5)P2 enables binding of domains such as ENTH in epsins that can begin to curve the membrane and assemble clathrin cages using their clathrin box motif and also attract additional adapter proteins via yet more SLiMs. In turn, additional sets of SLiM-bearing proteins stimulate the actin filament formation and attachment, necessary to fold and pull the invagination into the cytosol. Later, dynamin binds directly to PI(4,5)P2 on the membrane to complete the scission process. Once in the cytosol, the clathrin-coated vesicles are soon dismantled and the contents are included into the early endosomes. [For recent reviews of the process, see (28–30).] Many viruses enter the cell via endocytosis, using many different cell surface receptors (31). Viruses such as HIV and hepatitis C virus depend on the recognition of more than one receptor for entry, but in many cases, the stoichiometry of receptor engagement is unknown. Coronaviruses can enter cells through different routes that include RME and cell–cell fusion (32). In the case of SARS-CoV, the main entry route is endocytic and depends on endosome acidification (33, 34). However, protease-mediated activation of the spike protein relieves the pH dependence of viral entry, indicating that acidification is not a requirement per se, but acts by inducing the endosomal cleavage of the spike protein required for viral fusion (35, 36). The spike protein is cleaved either by the transmembrane protease serine 2 (TMPRSS2) at the cell surface or by cathepsin L within endosomes (37). The same entry route and proteases are used by SARS-CoV-2, and the use of endocytosis inhibitors indicates that the main entry route also seems to be endocytic (4, 38).

Autophagy is an evolutionarily conserved process in eukaryotes with multiple cellular roles that include the regulation of cellular homeostasis through the catabolism of cell components, immune development, and the host cell response to infection through pathogen phagocytosis (39). Viruses have evolved mechanisms to block the host cell antiviral response and can further hijack autophagy components to promote their survival and replication. This can be done through viral mimicry of host proteins coordinating autophagy or through the direct inhibition of the host autophagy machinery (40). Coronaviruses exploit the autophagy machinery through different mechanisms (41, 42). For example, MERS-CoV targets the BECN1 autophagy regulator for degradation, blocking the fusion of autophagosomes and lysosomes and protecting the virus from degradation (43). Coronaviruses repurpose cellular membranes to create double-membrane vesicles (DMVs) onto which the replication-transcription complex (RTC) is assembled, a process that involves recruitment of multiple autophagy components (41, 44, 45). DMVs in SARS-CoV-2 confine viral double-stranded RNA (dsRNA) concealing the viral genome from the innate immune system (46). Betacoronavirus mouse hepatitis virus (MHV) RTCs assemble by recruiting LC3-I, a nonlipidated form of the autophagy-associated protein LC3 (microtubule-associated protein 1A/1B–light chain 3) (41, 47), and SARS-CoV RTCs also colocalize with LC3 (44). Proximity-based mass spectrometry on the MHV replication complex further revealed that the RTC environment repurposes components from the host autophagy, vesicular trafficking, and translation machineries (45).

In the present work, we identify a set of conserved SLiM candidates in the ACE2 and integrin proteins, which are likely to act in the cell entry system of SARS-CoV-2 and provide molecular links to understand how the virus recognizes target membranes, enters into cells, and repurposes intracellular membrane components to drive its replication. These molecular links might provide previously unidentified clues toward drugging SARS-CoV-2 infections. We first focus on the extracellular SLiMs, before moving across the membrane to examine the cytosolic potential of the receptor tails. In a coincidently published paper, experimental testing of several motifs in the receptor tails is presented (48).

RESULTS
Extracellular receptor interplay and viral hijacking in the ACE2/integrin system
The identified RGD motif in the spike protein marks integrins as candidates for acting as co-receptors for SARS-CoV-2 entry. However,
Similarly to most SLiMs, the integrin-binding RGD motif has a low sequence information content, and the chance of random occurrence in protein sequences is relatively high. Therefore, the mere presence of an RGD motif in a sequence is not a strong indication of actual integrin binding. However, there are several features that make the spike-integrin interaction via the RGD motif plausible, including sequence- and structure-level information, gene expression profiles, the presence of accessory motifs, and protein-protein interactions. In the next sections, we review how this information gives credibility to the functional nature of the spike protein RGD as an integrin-binding motif and, more generally, to the existence of integrin hijacking by SARS-CoV-2.

The evolution of integrin-binding motif candidates within RBDs in the spike protein highlights that while the RGD motif is not conserved, the integrin-binding capacity might have evolved convergently in several betacoronaviruses. Owing to the high rate of recombination in coronaviruses (49), it is challenging to build proper phylogenies to trace their evolution. However, simply aligning homologs of the RBD from the Betacoronavirus genus (Fig. 1A) already shows that the RGD motif candidate is located in a locally less conserved region, hinting at the rapid evolvability of the site. The closest known homolog of SARS-CoV-2 is the RatG13 bat coronavirus containing TGD instead of RGD, which is incompatible with integrin binding. However, while the RGD motif itself is not conserved, several other members of the Betacoronavirus genus harbor other possible integrin-binding motifs. SARS-CoV and several of its close homologs, such as BM48-31/BGR/2008, contain KGD at this site. KGD can bind integrin as part of disintegrin binding, such as in the snake venom barbourin (50), but because disintegrins lacking KGD also bind integrin (51), and there is no evidence of KGD binding independent of disintegrins, we think that SARS-CoV KGD is less likely to be an active integrin ligand.

Considering more distant homologs of SARS-CoV-2, it becomes evident that the presence of an RGD/KGD site is not a universal feature of betacoronaviruses. The RBD of a moderately related Rousettus bat coronavirus does not contain any of the three residues of the RGD (Fig. 1B). However, other even more distant coronavirus sequences show a different potential integrin targeting motif at the same site. OC43 is a betacoronavirus that is one of the pathogens causing the common cold. Several OC43 RBD sequences show an NGR motif in nearly the same position as the SARS-CoV-2 RGD. NGR is an integrin interaction motif that becomes active upon the nonenzymatic natural deamidation of the asparagine residue preceding a glycine to isoaspartic acid, forming an ε-isoDGR site, which can recognize several α-integrins, as well as integrin α5β1 (52). The parallel evolutionary emergence of potential integrin-binding motifs at this location indicates that, despite the lack of conservation at the site, the SARS-CoV-2 RGD motif might be functional.

Normally, the functional importance of a protein region correlates with its conservation. Checking for sequence variances in the SARS-CoV-2 spike protein RGD motif across isolates showed that all 8841 (when checked on 9 June 2020) high-quality full spike protein sequences in GISAID (Global Initiative on Sharing Avian

---

**Fig. 1. The RGD motif of the SARS-CoV-2 spike protein.** (A) Multiple sequence alignment of a part of the SARS-CoV-2 spike RBD region using homologous sequences from betacoronaviruses of various evolutionary distances and showing the location of potential integrin-binding motifs in black. Virus names together with the host organisms, UniProt accessions (*or GenBank accession in the case of RatG13), and sequence region numberings are shown on the left side of the alignment. The location of the region shown in the alignment is indicated in a representative diagram of the spike protein, together with the location of the RGD motif and the region responsible for ACE2 binding. (B) Neighbor-joining tree of the multiple sequence alignment, with this particular set of sequences containing the potential high affinity, low affinity, and reverse integrin-binding motifs (RGD, KGD, and NGR) shown in red, orange, and green boxes, respectively. Only the sequence regions shown in (A) were used in the calculation of the tree. (C) Structure of the SARS-CoV-2 RBD as seen in the ACE2-bound form (PDB:6m17). The RGD motif is shown in red sticks. Regions in direct contact with ACE2 are shown in blue. Residues with missing atomic coordinates (indicating flexibility) in the unbound trimeric spike protein structures (PDB:6vsb, 6vxx, and 6vyb) are shown in transparency. Alignment and tree were prepared in Jalview (226) with Clustal colors. Structure was visualized using UCSF Chimera (228).
The sequence and structure context of the RGD motif can indicate possible target integrins. RGD motifs are recognized by several integrins, and specificity is determined mostly by the flanking residues of the core motif. As evidenced by crystallized integrin dimer-ligand complexes, the residue preceding RGD is in contact with the α subunit, whereas the residue after the core motif interacts with the β subunit. The immediate context of the SARS-CoV-2 RGD motif is 402-IRGDE-406 (Fig. 1A), which can give an indication about possible integrin targets. IRGD can be found in several native integrin-binding partners, including FREM1 (62), MFAP4 (63), and IGFBP1/2 (64, 65). These extracellular matrix proteins target integrins with αv, α5, and α9 subunits. RGDE is present in the native human integrin ligands TGFβ1, osteonectin, collagen α1(VI) chain, PSBG-9, and polydom, and in vivo and in vitro binding studies of the specificity profiles of these proteins (66–71) highlighted a post-RGD Glu to be efficient in binding to β1, β2, and β3 integrin subunits. Correlating these preferences with possible α- and β-integrin subunit pairings points to the most likely candidate target integrins for SARS-CoV-2 being αvβ1, αvβ3, αβ1, and αββ3. In vivo and in vitro integrin-binding studies have indicated that various α and αβ integrins share a large overlap in binding specificity for ligands, and therefore, any of these integrins might play a role in SARS-CoV-2 cell attachment and infection.

Most RGD-binding integrin dimers recognize the partner RGD motif in a long loop conformation that fits into the deep binding pocket of the receptor (Fig. S2A), including the integrin candidates identified by the RGD-flanking residues. However, available structures highlight that αvβ3 integrins have a different structural preference in their ligands. In this binding mode, the ligand is only in contact with the integrin α subunit via the Arg residue of the RGD motif. Therefore, the α subunit plays little role in specific ligand recognition. In contrast, the region following the RGD motif adopts an α helix and binds to the β-integrin subunit (Fig. S2B). In most known cases, this interaction is stabilized by two small hydrophobic residues fitting into two hydrophobic pockets on the surface of integrin β, establishing contacts with the three specificity-determining loops (72), conforming to a pattern of xRGDxxφ, where φ indicates a hydrophobic residue and x indicates any residue. This binding mode is known to be used by the growth factors transforming growth factor-β (TGF-β1) and TGF-β3 (72), and it is also mimicked by the cell attachment loop of the FMDV for cell entry (73). In its unbound state, the RGD motif of the SARS-CoV-2 spike protein RBD resides in a loop, followed by a helical structure containing two small hydrophobic residues, reminiscent of bound structures of αβ ligands (Fig. S2C). While the RBD is stabilized via three disulfide bridges, the RGD motif-containing region is on the far side of the domain. In addition, this region—together with the ACE2 binding site—has the highest average B-factor of the whole spike protein trimer (Fig. S2D), hinting at a possible structural rearrangement to accommodate the binding.

A major difference between TGF-β-type ligand and the RBD sequence is that RBD contains an extra residue between the RGD and the two hydrophobics, conforming to a pattern of RGDXqxxφ instead. On the basis of current knowledge, it is unclear how this would influence integrin binding; however, there are known αβ ligands that also deviate from the TGF-β subtype. Fibrin III-1 contains an integrin-binding region with the sequence RGDNKD-TACSNI, and it is a known ligand for integrins αβ1, αβ3, and αβ5 (74). The deviation from the canonical TGF-β-type motif is possibly a compromise between the—hitherto undescribed—specificity...
determinants of the three integrins, resulting in binding to several receptors with reduced affinity.

Motif-domain interactions are typically under heavy spatio-temporal regulation. Hence, the SARS-CoV-2 RBD-integrin binding can only occur if the possible target integrins are expressed on the infected host cells. Integrins \( \alpha_5\beta_1 \) (75) and \( \alpha_5\beta_3 \) (76–78), at least, have been observed in lung epithelial cells—the primary cells of infection in the lung—and are implicated in the emergence and progression of various diseases, including emphysema, non–small cell lung cancer, and mechanical injury of the lungs (79). SARS-CoV-2 infection has been observed to cause damage in various other tissues as well, including the heart, blood vessels, liver, and kidney (80). \( \alpha_5 \) integrins are near ubiquitous in major human tissues (81) and have been observed in all organs with observed damage from SARS-CoV-2 infections.

There are several other factors that point to an interplay between ACE2 and various integrins under normal cellular conditions. It has been shown that in heart tissues, ACE2 is able to bind the \( \beta_1 \) and \( \alpha_5 \) subunits of integrins in an RGD-independent manner, enhancing cell adhesion and regulating integrin signalling via the focal adhesion kinase (FAK) (60). It is unclear whether ACE2 interacts with integrins from the same cell, suppressing integrins by locking them in an inactive conformation, or adherent cells, acting as a direct inhibitor of integrins. However, the functional link indicates that integrins and ACE2 are expressed on the surface of the same cells in certain tissues, further corroborated by large-scale expression data (81). Furthermore, the RGD independence of the interaction means that while ACE2 and integrins are in complex, the RGD-binding site of the integrin is unoccupied, leaving it available for a potential interaction with a spike protein trimer.

Apart from the known interplay between ACE2 and integrins, there are additional features that indicate an even tighter cross-talk between the two receptors. RGD-mediated interaction to integrins and have been observed in all organs with observed damage from SARS-CoV-2 infections.

There are several other factors that point to an interplay between ACE2 and various integrins under normal cellular conditions. It has been shown that in heart tissues, ACE2 is able to bind the \( \beta_1 \) and \( \alpha_5 \) subunits of integrins in an RGD-independent manner, enhancing cell adhesion and regulating integrin signalling via the focal adhesion kinase (FAK) (60). It is unclear whether ACE2 interacts with integrins from the same cell, suppressing integrins by locking them in an inactive conformation, or adherent cells, acting as a direct inhibitor of integrins. However, the functional link indicates that integrins and ACE2 are expressed on the surface of the same cells in certain tissues, further corroborated by large-scale expression data (81). Furthermore, the RGD independence of the interaction means that while ACE2 and integrins are in complex, the RGD-binding site of the integrin is unoccupied, leaving it available for a potential interaction with a spike protein trimer.

Apart from the known interplay between ACE2 and integrins, there are additional features that indicate an even tighter cross-talk between the two receptors. RGD-mediated interaction to integrins and have been observed in all organs with observed damage from SARS-CoV-2 infections.

There are several other factors that point to an interplay between ACE2 and various integrins under normal cellular conditions. It has been shown that in heart tissues, ACE2 is able to bind the \( \beta_1 \) and \( \alpha_5 \) subunits of integrins in an RGD-independent manner, enhancing cell adhesion and regulating integrin signalling via the focal adhesion kinase (FAK) (60). It is unclear whether ACE2 interacts with integrins from the same cell, suppressing integrins by locking them in an inactive conformation, or adherent cells, acting as a direct inhibitor of integrins. However, the functional link indicates that integrins and ACE2 are expressed on the surface of the same cells in certain tissues, further corroborated by large-scale expression data (81). Furthermore, the RGD independence of the interaction means that while ACE2 and integrins are in complex, the RGD-binding site of the integrin is unoccupied, leaving it available for a potential interaction with a spike protein trimer.

Apart from the known interplay between ACE2 and integrins, there are additional features that indicate an even tighter cross-talk between the two receptors. RGD-mediated interaction to integrins and have been observed in all organs with observed damage from SARS-CoV-2 infections.

There are several other factors that point to an interplay between ACE2 and various integrins under normal cellular conditions. It has been shown that in heart tissues, ACE2 is able to bind the \( \beta_1 \) and \( \alpha_5 \) subunits of integrins in an RGD-independent manner, enhancing cell adhesion and regulating integrin signalling via the focal adhesion kinase (FAK) (60). It is unclear whether ACE2 interacts with integrins from the same cell, suppressing integrins by locking them in an inactive conformation, or adherent cells, acting as a direct inhibitor of integrins. However, the functional link indicates that integrins and ACE2 are expressed on the surface of the same cells in certain tissues, further corroborated by large-scale expression data (81). Furthermore, the RGD independence of the interaction means that while ACE2 and integrins are in complex, the RGD-binding site of the integrin is unoccupied, leaving it available for a potential interaction with a spike protein trimer.
motifs present in the human ACE2 tail, although it is >500 million years since their lineages diverged (102). In addition to the strong evolutionary conservation of these candidate motifs, their functional contexts are also biologically coherent, involving signaling by tyrosine kinases, endocytosis, autophagy, and actin filament induction (Table 1). In the following subsections, we briefly summarize each of the conserved motifs and their possible role in the viral entry mechanism.

The ACE2 tail contains a candidate YxxPhi endocytic sorting signal. The YxxPhi motif binds the μ2 subunit (UniProt: AP2M1_HUMAN) of the endocytosis AP2 adaptors by β-association (106). It is found in numerous cell surface receptors that have intrinsically disordered C-terminal tails (107). A small selection is listed in the database entry ELM:TRG_ENDOCYTIC_2, and while the motif has not been validated in ACE2, it is highly conserved (Fig. 3). When the Tyr is phosphorylated, this motif becomes an SH2-binding site, while in the apo form, it binds the μ2 adapter. Therefore, this motif can operate as a molecular switch. The residue following the Tyr makes a β-strand interaction and therefore cannot be a proline (PDB:1bxx). The phi position requires a bulky hydrophobic residue. The motif pattern can be represented by the regular expression [DENQSTAGYFP] defined in the ELM database for the SH2 domain–binding motif (Fig. 3) that is created upon phosphorylation of Tyr781. SH2-binding motifs are characterized by an invariant phosphotyrosine (pY) that is created following tyrosine kinase activation and allows binding to more than 100 types of SH2 domains present in human proteins (108). The pY residue is accompanied by additional binding determinants that frequently involve hydrophobic residues at the pY + 3 position, but can also involve other combinations, such as Asn at pY + 2 in Grb2-specific SH2 motifs or hydrophobic residues at pY + 4 in STAP-1 SH2 motifs (112; 110). Most SH2 motifs are also characterized by the exclusion of residues at certain positions following the pY, and in general, SH2-binding motifs show a high degree of cross-specificity (112) (109), limiting the power of bioinformatics predictions.

Cell culture infection assays with different coronaviruses, including SARS-CoV and SARS-CoV-2, have shown susceptibility to tyrosine kinase inhibitors, indicating the involvement of host tyrosine phosphorylation (25; 26; 27). The sequence found in ACE2 (781–YASID–785) matched the regular expression (Y) [DENQSTAGYFP] or [DENQSTAGYFP] defined in the ELM database for the SH2 domain present in NCK1/2 proteins, which belong to the class IA SH2 domains (110). Table 1. Known and predicted SLiMs in SARS-CoV-2 host-entry interactions. Previously identified motifs are marked with (✓). Regular expressions follow POSIX definitions (23). The symbols ‘x’ and ‘.’ mark any residues in the definition of main residues and regular expressions.

### Table 1. Known and predicted SLiMs in SARS-CoV-2 host-entry interactions.

<table>
<thead>
<tr>
<th>Region</th>
<th>Protein (UniProt accession)</th>
<th>Motif</th>
<th>ELM class*</th>
<th>Main residues</th>
<th>Regular expression</th>
<th>Start</th>
<th>End</th>
<th>Sequence†</th>
<th>Binding domain‡</th>
<th>Interaction partner§</th>
<th>Interaction type</th>
</tr>
</thead>
<tbody>
<tr>
<td>SARS-CoV-2 spike protein</td>
<td></td>
<td>RGD</td>
<td>LIG_RGD</td>
<td>RGD</td>
<td>RGD</td>
<td>403</td>
<td>405</td>
<td>RGD</td>
<td>PF00362 and PF01839</td>
<td>RGD-binding integrins, most probably ε2, ε3, and ε3,12</td>
<td>Host:Virus</td>
</tr>
<tr>
<td>Multibasic cleavage sites</td>
<td></td>
<td>–</td>
<td>RrrR</td>
<td>–</td>
<td>682</td>
<td>687</td>
<td>RRAR</td>
<td>SV</td>
<td>PF00082 or IPR001254</td>
<td>Furin-like PCs/ TMPRSS2</td>
<td>Host:Virus</td>
</tr>
<tr>
<td>CendR</td>
<td></td>
<td>LIG_NRP_CendR_1 RxxR [RK][O2][RS]</td>
<td>682</td>
<td>685</td>
<td>RRAR</td>
<td>PF00754</td>
<td>Neuruplin-1</td>
<td>Host:Virus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrin α</td>
<td>similar for other α chains</td>
<td>Multibasic cleavage sites</td>
<td>–</td>
<td>xKR</td>
<td>–</td>
<td>888</td>
<td>892</td>
<td>TKR</td>
<td>DL</td>
<td>PF00082</td>
<td>Furin-like PCs</td>
</tr>
<tr>
<td>Extracellular</td>
<td></td>
<td>Multibasic cleavage sites</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrin β</td>
<td>similar for other β chains</td>
<td>MIDAS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Furin</td>
<td></td>
<td>RGD</td>
<td>LIG_RGD</td>
<td>RGD</td>
<td>RGD</td>
<td>498</td>
<td>500</td>
<td>RGD</td>
<td>PF00362 and PF01839</td>
<td>Possibly RGD-binding integrin dimers</td>
<td>Host</td>
</tr>
<tr>
<td>MIDAS</td>
<td></td>
<td>–</td>
<td>DxsS</td>
<td>D,(TS),S</td>
<td>145</td>
<td>149</td>
<td>DLSSY</td>
<td>–</td>
<td>The acidic part of RGD-like ligands</td>
<td>Host</td>
<td></td>
</tr>
<tr>
<td>Multibasic cleavage site</td>
<td></td>
<td>–</td>
<td>R</td>
<td>–</td>
<td>697</td>
<td>716</td>
<td>RTEVEKAIRM SRSRINDAFR</td>
<td>IPR001254</td>
<td>TMPRSS2</td>
<td>Host</td>
<td></td>
</tr>
</tbody>
</table>

*Main present in NCK1/2 proteins, which belong to the class IA SH2 domains (110). No other SH2 entry catalogued in ELM matched the
<table>
<thead>
<tr>
<th>Region</th>
<th>Protein (UniProt accession)</th>
<th>Motif</th>
<th>ELM class*</th>
<th>Main residues</th>
<th>Regular expression</th>
<th>Start</th>
<th>End</th>
<th>Sequence†</th>
<th>Binding domain‡</th>
<th>Interaction partner§</th>
<th>Interaction type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intracellular</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrin β1 (P05106)</td>
<td></td>
<td>LIG_PTB, Apo_2</td>
<td>Nxx(FY)</td>
<td></td>
<td>E<a href="0.2">DST</a>[WFY]</td>
<td>767</td>
<td>773</td>
<td>TANNPLY</td>
<td>PF00373, PF00630</td>
<td>Talins (high affinity)</td>
<td>Kindlin-Filamin-A (binding to both apoPTB motifs simultaneously)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(L^RKP)(^PG)</td>
<td>767</td>
<td>773</td>
<td>TANNPLY</td>
<td>PF00640, PF02174</td>
<td>Talins (low affinity)</td>
<td>Kindlin-Filamin-A (binding to both apoPTB motifs simultaneously)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="0.4">LVMF</a>[LVMF]</td>
<td>777</td>
<td>783</td>
<td>TSTFTNI</td>
<td>PF02991</td>
<td>Shc (binding to both PTB motifs simultaneously)</td>
<td>Shc (binding to both apoPTB motifs simultaneously)</td>
</tr>
<tr>
<td>LIR autophagy</td>
<td>LIG_LIR, Gen_1</td>
<td>ExxYxxΦΦΦΦΦΦ</td>
<td></td>
<td></td>
<td><a href="0.2">EDST</a>[WFY]</td>
<td>777</td>
<td>783</td>
<td>TSTFTNI</td>
<td>PF0373, PF010480, PF00630</td>
<td>Talins (high affinity)</td>
<td>Kindlin-Filamin-A (binding to both apoPTB motifs simultaneously)</td>
</tr>
<tr>
<td>PTB (✓)</td>
<td>LIG_PTB, Phospho_1</td>
<td>Nxx(Y)</td>
<td></td>
<td></td>
<td>E<a href="0.2">DST</a>[WFY]</td>
<td>777</td>
<td>784</td>
<td>TGENPIY</td>
<td>PF0373, PF010480, PF00630</td>
<td>Talins (low affinity)</td>
<td>Kindlin-Filamin-A (binding to both apoPTB motifs simultaneously)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(L^RKP)(^PG)</td>
<td>777</td>
<td>784</td>
<td>TGENPIY</td>
<td>PF010480, PF00640, PF02174</td>
<td>Talins (low affinity)</td>
<td>Kindlin-Filamin-A (binding to both apoPTB motifs simultaneously)</td>
</tr>
<tr>
<td></td>
<td>ApoPTB (✓)</td>
<td>LIG_PTB, Apo_2</td>
<td>Nxx(FY)</td>
<td></td>
<td>L^RKP)(^PG)</td>
<td>777</td>
<td>796</td>
<td>TVVPNKY</td>
<td>PF0373, PF010480, PF00630</td>
<td>Talins (low affinity)</td>
<td>Kindlin-Filamin-A (binding to both apoPTB motifs simultaneously)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="0.4">LVMF</a>[LVMF]</td>
<td>777</td>
<td>796</td>
<td>TVVPNKY</td>
<td>PF010480, PF00640, PF02174</td>
<td>Talins (low affinity)</td>
<td>Kindlin-Filamin-A (binding to both apoPTB motifs simultaneously)</td>
</tr>
</tbody>
</table>

*Motif identifier as in the ELM resource. †"|" denotes cleavage points for protease-recognition motifs. §PC, proprotein convertases. □Not a SLiM but a structural motif. □□Defined through use of Pfam (103) or InterPro (104), where applicable. □□□
tail. Proteins known to contain this motif are listed in entry ELM:LIG_SH2_NCK1_1. We have since learned that an ACE2 phosphorylated Tyr 781 (pTyr 781) tail peptide does not bind to NCK1 (48). Upon re-examination of the SPOT arrays in (111,112), we noted that the strong preference at pY + 3 is for Val and Pro. While Ile is tolerated at pY + 3 in the context of the high-affinity EPEC Tir (enteropathogenic Escherichia coli translocated intimin receptor) sequence (111), it is not tolerated in the context of random peptide pools (112). This would indicate that NCK can only tolerate a weak Ile residue at pY + 3 when a strong residue such as Glu and Asp is found at pY + 1, such as Asp in EPEC Tir. The presence of the weak aliphatic residue Ala at pY + 1 in ACE2 would explain the lack of binding for the ACE2 tail motif. This evidence indicates that the ELM pattern needs correcting to allow only one weak amino acid at either of pY + 1 or pY + 3 in the regular expression.

Other class 1A SH2 domains with a strong preference for Ile at the +3 position in SPOT array include the SH2 domains of the SRC family kinases (SFKs). A regular expression for SRC family SH2 domains allowing for weak/strong residues +1 and +3 positions and compatible with the SPOT arrays could be ((Y)[DE][KRHG][DESTAPILVMFYW][^KR])|(Y)[NQSTAILVMFY][KRHG][ILV]^KR) (Table 1). This pattern matches the ACE2 tail. The ACE2 YASID sequence has a weak Ala at pY + 1 in ACE2 would explain the lack of binding for the ACE2 tail motif. This evidence indicates that the ELM pattern needs correcting to allow only one weak amino acid at either of pY + 1 or pY + 3 in the regular expression.

The residues present at pY + 1, pY + 2, and pY + 4 should rule out that the ACE2 YASID motif can be a strong Grb2, CRK, and STAP-1 SH2 domain binder, and binding to SH2 domains in the transcription factors signal transducer and activator of transcription 1 (STAT1), STAT3, and STAT5 is also unlikely due to the lack of adequate specificity determinants. However, other SH2 domains, particularly ones with low observed specificity (e.g., PTPN11_N, PLCgamma1_C, and SH2D1A), could be recruited by ACE2 when there is coexpression in the same cell type. Experimental validation will be required to test these hypotheses.

Tyr 781 in ACE2 also overlaps with a candidate phosphorylation-independent NPY IBAR-binding motif (ELM:LIG_IBAR_NPY_1). This motif was initially described in the bacterial secreted protein...
Tir from pathogenic strains of *Escherichia coli*, such as enterohaemorrhagic *E. coli* (EHEC). The NPY tripeptide recognizes and binds with a 60 μM affinity to inverse Bin-Amphiphysin-Rvs (I-BAR) domains in adaptor proteins like insulin receptor substrate protein of 53 kDa (IRSp53) and its homolog insulin receptor tyrosine kinase substrate (IRTKS) (117, 118). I-BAR domains bind to the plasma membrane to favor weak membrane protrusions, and the preference of I-BAR domains for negative membrane curvatures enables a positive feedback loop that can result in the formation of lamellipodia, filopodia, and other types of membrane protrusions (111). The NPY motif in the cytosolic region following the transmembrane helix (112–114). During EHEC infection, the bacteria use the NPY motif as a scaffold to localize the injected bacterial protein EspFU to viral infection (126). Ezrin is a protein involved in cell morphology and apical membrane remodeling that acts as a membrane-cytoskeleton linker. Ezrin recruits F-actin through its C-terminal domain and can also bind to IRSp53 located at negatively curved membranes (127, 128), suggesting that while the NPY motif acts at earlier stages of viral attachment, the spike protein–Ezrin interaction might work during or after viral fusion, to promote the recruitment of actin-regulatory components to viral fusion sites.

Apart from the endocytic sorting signal, the SH2 binding, and the I-BAR-binding motif, Tyr<sup>120</sup> is also part of an LC3-interacting region (LIR) autophagy motif candidate (Fig. 3). Autophagy, the recycling of cellular material, is vital for cellular homeostasis. Many pathogens must control the autophagy response to establish productive

infection (39). It has been shown that coronaviruses, including those that infect humans, subvert autophagy components to promote viral replication at DMVs associated to the RTC (43, 47, 129, 130). The LIR motif is required for the interaction of a target protein with autophagy-related protein Atg8 in yeast, or its homologs LC3 and GABARAP in human, to facilitate autophagy of the target via the autophagosome (131). The LIR motif has been catalogued in the ELM resource entry ELM:LIG_LIR_Gen_1, and ELM detected a candidate motif in the human ACE2 cytosolic tail sequence (Fig. 3). After the LIR motif was annotated in ELM, a more recently solved LC3-LIR structure (PDB:5cx3) showed that the interacting peptide is longer, with one or two additional hydrophobic interactions (132). LIR enters a hydrophobic groove bordered by positively charged residues. A core [WFY]xx[ILMV] enters the deepest part of the groove. On either side of the core, the interacting residues can be flexibly spaced. The core must be preceded by a negatively charged residue (which might be enabled by phosphorylation). Furthermore, the motif core is followed by a flexibly spaced hydrophobic residue. There is often a negatively charged residue preceding this hydrophobic position: It can make favorable interactions with counter charges but is not an absolute requirement, so is not included in the revised motif pattern. On the basis of the structure (PDB:5cx3) and some SPOT arrays (132–134), the updated regular expression [EDST][0,2][WFY][^RKP][^PG][ILMV][0,4][LIVFM] matches the motif instances annotated in ELM. This revised motif is conserved in the mammalian ACE2 cytosolic tail as well as hagfish and ghost shark, but not in birds, reptiles, or bony fish. The ACE2 LIR motif candidate can potentially enable the incoming coronavirus to attract autophagy elements such as LC3 to the structures where the virus replicates and assembles. In line with this, a nonlipidated form of the LC3 protein has been shown to be associated with the RTCs of MHV and SARS-CoV (41, 44, 47). This brings up the interesting possibility that ACE2 remains associated with the membranous structures where SARS-CoV-2 replicates at later infection stages, assisting in the repurposing of autophagy components required for viral replication. Technical issues hampered the comprehensive testing of phosphorylated ACE2 peptide sequences containing the LIR candidate, but the nonphosphorylated peptide did not show meaningful binding (48). However, phosphorylation of Ser783 seems to induce a weak binding with MAP1LC3A and GABARAPL2 domains, albeit with affinities not reaching physiological relevance (48). So far, the evidence is not enough to support LIR functionality but perhaps multiphosphorylation and/or a longer tail sequence could deliver a stronger interaction.

The ACE2 tail region C-terminal to the overlapping motifs centered around Tyr781 contains two additional motif candidates. The first such candidate is an apoPTB domain-binding motif. Certain members of the large PTB domain family were initially discovered to bind to phosphorylated NPxY motifs, hence the designation “phospho-tyrosine binding domain” (135). The NPxY motifs in cytosolic tails of receptors, including integrins, are regarded as endocytosis sorting signals (107). It was later discovered that PTB domains in the endocytic internalization adapter protein Dab1 could also bind nonphosphorylated Nxx[FY] motifs (apoPTB motif) and that this might be the case for the majority of PTBs (136). Representative receptors with apoPTB motifs are in the database entry ELM:LIG_PTB_Apo_2. The core Nxx[FY] motif is conserved in all the vertebrate ACE2s (Fig. 3). For the Dab1 endocytic adapter class of apoPTB motifs, there is a hydrophobic requirement two residues before the Asn. In ACE2 of fishes such as the hagfish and coelacanth (Latimeria chalumnae), the residue is hydrophobic (Fig. 3), suggesting that this motif is present. However, in most other species including human, Glu predominates at this position: Therefore, if this notably conserved Nxx[FY] is an apoPTB motif, it should then bind a PTB protein other than the Dab1 class. The apoPTB motif binds as a short β-strand (β-augmentation) followed by a β-turn. Proline is rejected at the first position of the motif, which is a strand-forming residue, and therefore, the minimal regular expression for this motif is [^P][N...][FY]. As with the phosphorylated versions, the apo-motifs are tightly connected to endocytosis (136). The conservation of this motif in the homologous position of the cytoplasmic chain of the partially collagen collectrin protein (UniProt: CLTRN_HUMAN; fig. S3) indicates that this motif instance has an even earlier evolutionary origin than the origin of ACE2 itself, hinting at a key role in internalization. As expected, because the specificity is not yet defined, Dab1 and four other tested PTB domains did not bind to the ACE2 tail region (48). A poorly soluble sorting nexin 17 (SNX17) FERM domain was found to bind with ≈100 μM affinity, providing an ambiguous result.

The very C-terminal region of ACE2 contains a TxFS PDZ-binding motif (PBM) candidate. Among other motif-binding modules, PDZ domains come in great abundance in human and other multicellular animals (137). PDZ domains take part in a variety of biological processes including cellular signaling and activity at the neuronal synapse (138). These domains bind by β-strand augmentation to SLIMs that are called PBMs, most commonly known to be found in the C terminus of fully or partially disordered proteins. These interactions are widely studied and their link to various diseases and infections has been previously established (139). A PBM candidate is also found in the very C terminus of the cytosolic tail of all vertebrate ACE2 proteins (Fig. 3). Motifs following a pattern [ST][ACVILF] are a common PBM variant, described in the ELM resource entry ELM:LIG_PDM_Gen_1. There are multiple functional examples of this motif. However, in the ACE2 protein, the matching sequence has not been characterized. Because the tail of ACE2 is facing the cytosol, it is available to interact with PDZ domains with the appropriate specificity (138).

Two PDZs in two different adapter proteins—Na(+)/H(+) exchange regulatory cofactor NHERF3 and SH3 and multiple ankyrin repeat domains protein 1 SHANK1—have been previously identified to be able to bind to TxFS sequences (140), which makes both of them candidates for an interaction with the ACE2 C terminus. NHERF3 is colocalized with ACE2 in intestinal tissue, and its PDZ domains were previously validated to interact with PBMs in transmembrane proteins on the cytosolic side of the membrane (141), so it is possible that they come in proximity with the ACE2 tail containing the TxFS motif and possibly bind it as a part of ion exchange regulation of small-molecule transport activities. NHERF3 is known for its involvement in sodium ion–dependent transporter activity (142), and ACE2 was also shown to interact with a sodium-dependent transporter (57), which could be one of the leads toward unraveling the possible interaction between NHERF3 and ACE2. Kliche et al. (48) confirmed ACE2 tail binding with good affinity for both NHERF3 and SHANK1. They also measured low micromolar affinity for the PDZ domain of SNX27, which is involved in retrograde transport from the endosome to the plasma membrane. Although plausible, whether or not NHERF3 and SNX27 are PDZ domain–containing
proteins interacting with ACE2 is an open question that will require follow-up experiments in the cell.

**Tyr<sup>781</sup> in the ACE2 tail creates a potential multiway molecular switch regulated via phosphorylation**

The tyrosine at residue 781 in ACE2 is a part of the motif patterns for four of the motifs listed above (Fig. 3 and Table 1) but must be phosphorylated to act as an SH2-binding motif. We searched the ACE2-related literature for reports of phosphorylation but were unable to find any with strong site identification. Examination of the human ACE2 entry in the database PhosphoSitePlus (143) revealed that high-throughput (HTP) phosphoproteomic studies, but no low-throughput (LTP) studies, identify pTyr<sup>781</sup>. Thirteen HTP measurements identified phosphorylation at Tyr<sup>781</sup>, and this residue is the only ACE2 phosphosite that is reproducible across multiple HTP datasets (Fig. 4). For example, pTyr<sup>781</sup> was one of 318 unique phosphopeptides belonging to 215 proteins analyzed from an erlotinib-treated breast cancer cell line model (144). Therefore, this site fulfills the phosphorylation requirement to be an SH2-binding motif.

As outlined above, four candidate sequence motifs overlap in the region surrounding Tyr<sup>781</sup>: the YxxPhi endocytic sorting signal (ELM:TRG_ENDOCYTIC_2), an SH2 motif that mediates binding to SFKs, an NPY I-BAR–binding motif (ELM:LIG_IBAR_NPY_1), and the LIR autophagy motif (ELM:LIG_LIR_Gen_1). While the YxxPhi, NPY, and LIR motifs require an unphosphorylated state of Tyr<sup>781</sup>, the SH2 motif requires Tyr<sup>781</sup> phosphorylation, creating the opportunity for a multiway phospho-switch acting in this region of ACE2 that directs different steps of the SARS-CoV-2 infection cycle. In support of this proposal, Kliche et al. (48) confirmed that the ACE2-YxxPhi interaction is negatively regulated by phosphorylation and that binding to the FYN SH2 domain requires Tyr<sup>781</sup> phosphorylation. The relative affinities of the ACE2 tail binders, which is still to be fully established, will dictate the competition between the interactions and the functional output. Current results indicate that the phosphorylated ACE2 tail can reach low micromolar affinity for SFKs and that the unphosphorylated state can bind to the AP2 µ2 subunit with moderate affinity, while physiologically relevant interactions with autophagy components and I-BAR domains are still to be demonstrated. The state of this switch could be controlled by protein localization and by tyrosine kinase activity involving SRC/Abl and other tyrosine kinases, which are known to have increased abundance during endosomal processes (115) and viral infection (18) including in coronaviruses (2, 25–27). Similar switches have been described before, as with the cytotoxic T lymphocyte–associated protein 4 (CTLA-4) receptors, where SRC tyrosine kinases dictate the binding preferences of overlapping YxxPhi and SH2-binding motifs. In the unphosphorylated state, endocytosis is favored, whereas T cell activation brings about Tyr phosphorylation, shutting down endosomal recycling and initiating signaling through the recruitment of SH2 domain–containing proteins (106, 145–148). The CagA effector from Helicobacter pylori provides an example of a multiway molecular phospho-switch, where the choice for senescence versus cell proliferation is dictated by the SH2 domain–containing protein that forms a complex with phosphorylated CagA (24). Additional regulation can create a temporal gradient of the phospho-signal: CagA leads to remodeling of the actin cytoskeleton through its sequential phosphorylation by tyrosine kinases. Initial phosphorylation by SRC creates a negative feedback loop that terminates SRC signaling through activation of the SRC inhibitor Csk in the early stages of infection, while phosphorylation by Abl kinases leads to concerted changes in the phosphorylation of actin-regulatory proteins that drive actin-cytoskeletal rearrangements at later time points of infection (149).
A similar temporal regulation could be at play in SARS-CoV-2 endocytosis. This might be enacted by a Tyr$^{61}$ phospho-switch. The early attachment phase could be characterized by unphosphorylated Tyr$^{61}$ that allows the YxxPhi and NPY motifs to be active. During this phase, the YxxPhi motif could initiate RME by binding the AP2 complex μ2 subunit, recruiting clathrin and other endocytic components to the viral attachment sites. In addition, some viruses can "surf" along filopodia by myosin-mediated actin cytoskeleton movements that transport the viral particles to the entry sites at the cell body, ultimately increasing their entry rate (125). The formation of these membrane protrusions could be promoted by the I-BAR–binding NPY motif. The relative affinity and availability of binders might dictate the sequential or concerted use of the YxxPhi and NPY motifs during the initial stage. Following the initial steps of membrane attachment and clathrin coat formation, actin polymerization is required to internalize the endocytic vesicles. This second step could be brought about by SFK-mediated Tyr$^{781}$ phosphorylation that leads to disengagement of the AP2 μ2 subunit and I-BAR–containing proteins and to activation of actin-regulatory proteins through SFK recruitment. SRC and Abl, two of the SFKs predicted to bind the SH2 motif, are known to promote RME and actin cytoskeletal rearrangements (113, 115).

An alternative scenario that is not mutually exclusive with temporal regulation might be enabled by the multimeric nature of the spike protein and by attachment of several viral particles to a membrane domain, leading to adjacent ACE2 tails on the intracellular side that expose both phosphorylated and unphosphorylated motifs, allowing these three signaling steps to take place simultaneously. The separation between the RBD-binding sites in the ACE2 dimer is 68 Å calculated from PDB: 6m17 (57), in close agreement with the separation between the YxxPhi binding sites in the ACE2 dimer (fig. S1). While the outward orientation of the RBD-binding sites in ACE2 might preclude stable contacts between two RBDs and an ACE2 dimer, the spatial proximity implies that both ACE2 subunits are likely activated by the dynamic interaction of a spike protein trimer with an ACE2 dimer. The presence of several parallel routes for the recruitment of cytoskeleton components involving the NPY and SH2 motifs could provide the robustness needed to ensure the actin reorganization required for the uptake of virus-containing vesicles into the cytosol. Following endocytosis and fusion, viral components are released into the cell and viral replication takes place. SFKs have been shown to be inactive in the endosomal compartments, which would lead to dephosphorylation of Tyr$^{781}$ following endocytosis (115). During this phase, the last component of the switch could come into play, when the ACE2 protein that remains bound to spike protein–coated membranes could promote the hijack of autophagy components necessary to assemble the viral replication factories. However, the functionality of the LIR motif has not yet been established and might require other PTMs of the ACE2 tail, as suggested by Kliche et al. (48).

**Known and candidate motifs in the β-integrin tails**

Integrin β tails are short cytosolic C-terminal intrinsically disordered regions, similar to the analyzed region of ACE2. The three most probable integrin β subunit candidates at play in SARS-CoV-2 viral entry are β3, β5, and β6. The C-terminal tails of all three subunits share a high degree of sequence similarity (with β3 and β6 being almost identical) and, similarly to ACE2, contain several known and candidate SLIMs (Table 1 and Fig. 5, A and B) that propagate signals in the cytoplasm and regulate integrin activity not only through intracellular pathways but also changing the structural state of the ectodomains determining ligand-binding capacity (150). In addition, all three integrin β tails are very highly conserved (figs. S4 to S6), hinting at their high functional importance.

Integrin β tails contain a highly charged patch in their membrane-proximal region (Fig. 5A). This region is indispensable for the interaction between integrins and tyrosine kinases, including the SRC kinase Fyn (151) and FAK, most probably via the direct interaction with paxillin (152). Through these interactions, integrins regulate cytoskeletal remodeling (153) and the promotion of cell survival (154), as well as regulation of FA assembly and cell protrusion formation (155). In turn, FAK regulates integrin recycling and endosomal trafficking (156, 157).

Now, there is no consensus sequence motif describing these interactions, although a definition of HDR[KR]E has been proposed (158), matching integrins β1, β3, β5, and β6. This motif is under heavy regulation by several mechanisms. First, the interaction with tyrosine kinases seems to involve additional residues N-terminal of the charged motif core—most notably, the conserved lysine preceding the hydrophobic patch (159)—that are only accessible in the active state of the integrin dimer, as these regions are buried in the membrane otherwise (160). Second, the D residue of the motif forms a salt bridge with the cytosolic tail of the α subunit of the integrin in the inactive conformation of the receptor. Thus, this motif region is dependent on integrin activation regulated by ligand binding and intracellular interactions mediated by the downstream NpxY motifs.

The tails of integrins β1, β3, and β6 contain two regions that match the apoPTB motif (Table 1 and Fig. 5A) as either NpxY (with two matches in integrin β1 and 1-1 matches in integrins β3 and β6) or qxnNxyx (with 1-1 matches in integrins β1 and β6). Furthermore, these regions are known to have Tyr phosphorylation, matching the phosphorylated motif definition as well (ELM:LIG_PTB_Phosho_1). These regions are known to be able to form β-turns and are recognition sites for PTB domains. In addition, NpxY motifs are the major sorting signals mediating interactions with FERM domains for regulating endosomal trafficking (161). In β-integrin tails, these motifs recruit adaptor proteins and clathrin, serving as sorting signals (162), and the NpxY motifs in the β1 tail have a direct connection to viral entry for reovirus (163).

The NpxY motif switches mediate several interactions. The membrane-proximal NpxY motif binds talin-1, serving as a connection between the plasma membrane and the major cytoskeletal structures (164). Considering the expression profiles of talins, the most likely interaction partner of lung-expressed integrins is talin-1. Talin-1 contains a FERM domain, similarly to Ezrin, which establishes a direct interaction with the SARS-CoV spike protein upon viral fusion (126). However, the interaction between the RBD and integrins offers the virus an earlier point of interference with the cytoskeletal system, being able to modulate it cooperatively with the ACE2 actin-regulatory elements (NPY and SH2 motifs) before and during cellular entry. The talin/integrin interaction, however, presents a feedback loop: The binding of talin on the cytoplasmic side induces a structural rearrangement on the ectodomains of integrins, enabling a higher affinity interaction with RGD motif–containing ligands (165).

The membrane-proximal NpxY motif is also a binding site for docking protein 1 (DOK1), a negative regulator of integrin activation. DOK1 is in direct competition with talin for binding integrins (165). The competition is fundamentally influenced by phosphorylation.
on Tyr\textsuperscript{783} (for integrin \(\beta_1\); fig. S7A), Tyr\textsuperscript{773} (for integrin \(\beta_3\); fig. S5B), and Tyr\textsuperscript{762} (for integrin \(\beta_6\); fig. S7B) of the NPxY motif. The unphosphorylated motif has a higher affinity toward talin, whereas phosphorylation prefers DOK1 (166); thus, the tyrosine acts as a phospho-switch regulating integrin activation.

The membrane-proximal NPxY motif also presents a binding site for a largely phosphorylation-independent interaction with the integrin cytoplasmic domain–associated protein-1 (ICAP-1). ICAP-1 is a fundamental regulator of the assembly of FAs and ICAP-1 knockdown reduced FA assembly (167), possibly working in conjunction with the membrane-proximal charged region. ICAP-1 seems to be specific for \(\beta_1\), and hence, the therapeutic considerations for targeting this pathway require the verification of the type of integrins expressed on AT2 cells (and other related cell types).

The membrane distal NPxY motif is a binding site for the FERM domain of kindlin (168). This interaction requires the integrin tail to be nonphosphorylated, and phosphorylation on Tyr\textsuperscript{795} (for integrin \(\beta_3\)) or Tyr\textsuperscript{785} (for integrin \(\beta_6\)) can switch off the interaction with kindlin-2 (169) (no corresponding Tyr phosphorylation has been identified in \(\beta_6\) tails as of yet). Kindlin binding (together with talin binding) is a crucial step in integrin activation and hence regulates the availability of integrins for extracellular ligands (170) and was also suggested to play a role in TGF-\(\beta_1\) signaling (171).

The two NPxY(-like) motifs in the integrin \(\beta_6\) tails not only constitute two separate phospho-switches (Fig. 5, fig. S7, and Table 1) but also act in synergy to give rise to more complex regulation. Filamin and the PTB domain region of Shc1 each bind to both NPxY motifs (172, 173). Shc is an adaptor protein playing a key role in mitogen-activated protein kinase (MAPK) and Ras signaling pathways, and its interaction with integrin \(\beta_3\) requires both phosphorylations on Tyr\textsuperscript{773} and Tyr\textsuperscript{785} (172, 174). In contrast, binding of the immunoglobulin domain of filamin-A requires both tyrosines to be in a non-phosphorylated state. The filamin-A interaction can be considered as a main shutdown switch in integrin signaling, as this interaction induces the closed conformation of the integrin ectodomains, decreasing the chance of ligand binding (173). In addition, binding partners using both NPxY motifs may also serve as stronger modulators of endosomal trafficking, switching on enhanced signals.

Integrins are known to be connected to autophagy regulation, and therefore, motif identification and analysis might help suggest possible underlying molecular mechanisms. The connection between autophagy and cell adhesion has already been described, showing that both reduced FAK signaling (175) and detachment from the extracellular matrix via integrins (176) enhance autophagy. Atg-deficient cells have enhanced migration properties, and at the molecular level, there seems to be a direct connection between Atg proteins and integrins as well: autophagy stimulation increases the colocalization of \(\beta_1\) integrin–containing vesicles with LC3-stained autophagic vacuoles, whereas autophagy inhibition decreases the degradation of internalized \(\beta_1\) integrins (177). In Drosophila cells, it has been shown that the Wiskott-Aldrich syndrome protein and SCAR homolog (WASH) plays a connecting role between integrin...
recycling and the efficiency of phagocytic and autophagic clearance (178). However, molecular details about how this connection is brought about are unclear.

Sequence analysis of integrin β3 and β6 tails shows a potential Atg-targeting LIR motif (Fig. 5A), similarly to the ACE2 tail. However, neither β-integrin tails conform to the regular expression introduced in earlier sections, as the hydrophobic residue following the core motif is a tyrosine (Tyr785 for β3 and Tyr774 for β6). Thus, to capture this instance as well, the regular expression needs to be modified to [EDST].[0,2][WFY][\^RKP][\^PG][ILMV].[0,4][LIVFM]. LTP phosphorylation assays have determined that both Tyr773 and Tyr785 for β3 are phosphorylated in live cells (Fig. 5B). However, such assays have also determined additional phosphorylation sites in the β3 tail, Thr777, Ser778, Thr779, and Thr784. These phosphorylations are not connected to the NPxY motif switches in any known way but could serve as charge-based switches for the LIR motif. The peptide binding assays presented in the accompanying paper by Kliche et al. (48) show that phosphorylations introduced in the N-terminal tandem sites yielded low micromolar binding affinities. In addition, phosphorylation of Tyr785 further increases affinity, showing that the loss of the favorable interaction mediated by the C-terminal hydrophobic residue can be well compensated for by electrostatic interactions. While the current motif definition does not exactly fit the β6 tail, there are also LTP phosphorylation assay data (179) for the existence of these phosphorylations in the corresponding residues, hinting at the possibility of the presence of a slightly modified motif. For β3, as well as for β1 and β5 tails, the phosphorylation provides the negative charge required upstream of the FxSfXY LIR motif hydrophobic core. Phosphopeptides spanning the candidate region should also reveal whether the LIR motif-like region is a functional Atg-binding site in integrin β1. Such experiments can also shed light on the existence of a rheostat-like behavior of multi-phosphorylation, already demonstrated to a certain extent for the β3 LIR. The motif found in integrin β3 is also present in integrin β5, and the motif candidate identified in integrin β1 is also present in integrin β6.

Potential synergy between the ACE2 and integrin intracellular motifs

Bringing together the candidate SLiMs identified in the integrin β and ACE2 tails potentially strengthens the functional links between them and provides an emergent picture of SLiM-driven cooperative switches driving viral attachment, entry, and replication (Fig. 6). Following attachment of the spike protein to the receptors, the two NPxY motifs in the integrin subunit could act cooperatively with the apoPTB and YxxPh motif of ACE2 as sorting signals that mediate the internalization of viral particles into endosomes. The presence of several endocytic motifs in close proximity would strengthen the interaction with the endocytosis apparatus, creating a high-avidity environment for recruitment of RME components (107). During this time, the phosphorylated integrin NPxY motifs would also reinforce viral attachment through inside-out signaling, stabilizing the integrin ectodomain in the open, high ligand affinity conformation. As discussed previously, RME also involves the recruitment of adaptor molecules that activate rearrangements of the actin cytoskeleton required for the internalization of the endocytic vesicle. At this stage, the NPY and SH2 motifs in ACE2 would recruit several molecules that mediate actin polymerization signaling, prominently I-BAR–containing proteins IRSp53 and IRTKS as well as actin cytoskeleton regulators activated by SFKs. While most of this actin signaling would serve to allow viral entry, additional actin recruitment processes could occur following viral fusion, such as that initiated by the interaction between the spike protein and Ezrin. Last, at later stages of infection, both integrins and ACE2 might remain attached to virus-associated DMVs and other replication-competent membranes where the RTC assembles. At this stage, ACE2 and integrins might cooperatively mediate the recruitment of autophagy components such as LC3 through the LIR motifs located in the cytosolic tails of both molecules.

SLiMs and their potential therapeutic implications

The analysis of candidate SLiMs in ACE2 and integrins suggests that SARS-CoV-2 hijacks both receptors, co-opting their SLiMs to drive viral attachment, entry, and replication. This creates an opportunity for drugging these interactions, or the processes they control, through host-directed therapies (HDTs) to prevent viral entry. On the basis of the identified candidate interactions, we collected a list of potentially useful drugs (Table 2) together with ChEMBL accessions (180); several are already registered for clinical trials (181).

The RGD sequence is used by a large number of viruses for cell attachment, via integrins (13). RGD mimics have been developed as inhibitors of integrin–extracellular matrix protein interaction for a variety of diseases. A cyclic RGD peptide [c-RGD(NMe)V, cilengitide] has been developed clinically for glioblastoma treatment and other cancers. It proved safe but did not enhance the survival benefit (182). SARS-CoV-2 has a unique RGD sequence in the vicinity of the ACE2 binding region of its spike protein. It has been proposed that integrins may have a potential role for infectivity (12). If so, RGD mimetics might be able to block the RGD-binding site(s) on target cells and block the attachment of the virus. Another application that has been suggested is bacterial sepsis (sepsis is also a dreaded complication in COVID-19 patients), and experimental evidence in animals is available (183). Cilengitide is relatively specific for integrin α5β1 but also active on αvβ3, αvβ5, αvβ6, α1β3, α2β1, α4β1, and α5β1 (in decreasing order of activity). The antibody abituzumab (DI-17E6) is a pan-αv antibody, meaning it is also active against other αv integrins and, consequently, may be better suited for blocking virus entry. It has been clinically tested in several cancer indications (184, 185).

As discussed above, tyrosine kinase–mediated phosphorylation plays an important role in virus entry and maturation, and several tyrosine kinase inhibitors have entered the clinic and some show effects on viral infection in cell culture. For example, saracatinib, an SRC and Abl inhibitor that has completed several clinical trials, mainly targeting cancers, inhibited replication of different coronaviruses including MERS-CoV, SARS-CoV, and HCoV-229E in cell culture infection experiments (27). After internalization and endosomal trafficking, imatinib, an Abl inhibitor, prevented fusion of SARS-CoV and MERS-CoV virions at the endosomal membrane in infected cell culture experiments (25). Using the avian model virus IBV, imatinib and two other Abl inhibitors (GNF2 and GNF5) prevented the fusion of the spike protein to the membrane of the target cell as well as cell-cell fusion and syncytia formation (2). More recently, tyrphostin A9, a platelet-derived growth factor receptor (PDGFR) tyrosine kinase inhibitor, came out from an HTP screening using cytopathic effect as readout and also showed in vitro inhibitory capacity to transmissible gastroenteritis virus (TGEV), an alphacoronavirus that infects pigs (26). The authors also showed...
that tyrphostin A9 has a broad antiviral spectrum, being active against three other tested coronaviruses: MHV in murine L929 cells, porcine epidemic diarrhea virus in primate Vero cells, and feline infectious peritonitis virus in feline CCL-94 cells. The mode of action was found to be through p38 MAPK, at the post-adsorption stage. As FAK has been implicated in viral entry for other viruses including influenza A (186), experimental drugs targeting FAK, including some in clinical trials (187), can be considered for studying the potential spike protein-induced integrin signaling. Now, 39 tyrosine kinase inhibitors are approved by the U.S. Food and Drug Administration (FDA): 11 target nonreceptor protein–tyrosine kinases and 28 inhibit receptor protein–tyrosine kinases (188). Consequently, tyrosine kinase inhibitors may be good candidates to test for their effect on SARS-CoV-2. For example, an inhibitor of the Abl and PDGFR kinases, flumatinib mesylate, showed 42% reduction of SARS-CoV-2 infection of Vero E6 cells at 2.5 μM (189). As part of the United Kingdom’s ACCORD (Accelerating COVID-19 Research & Development) program, a clinical trial is underway to evaluate bempidainib, a specific inhibitor of the receptor tyrosine kinase AXL in COVID-19 (190). AXL acts as a pleiotropic inhibitor of innate immunity (191) and is also a receptor for Ebola virus (192).

A number of protease inhibitors are now discussed for SARS-CoV-2 treatment. Serine protease inhibitor camostat mesylate is active against TMPRSS2 and blocks cell entry (4). Nafamostat mesylate—originally developed as a trypsin inhibitor (193)—also has been shown to inhibit TMPRSS2. Nafamostat mesylate is an approved anticoagulant in Japan, with clinical testing for COVID-19 infections now being conducted. The spike protein of SARS-CoV-2 contains a furin cleavage sequence (PRRARS|V). Consequently, furin convertase inhibitors are considered as antiviral agents (194). A prime example of such inhibitors is decanoyl-RVKR-CMK, which has been shown to inhibit cleavage of the SARS-CoV-2 spike protein at the S1/S2 site by furin (90). A large drug screen identified four drugs that targeted host cysteine proteases in SARS-CoV-2–infected human cells including VBY-825 (cathepsin B/L), ZLVG CHN2, ONO 5334 (cathepsin K), and MDL-28170 (cathepsin B and calpain I/II), with the latter two inhibiting SARS-CoV-2 replication in human induced pluripotent stem cell (iPSC) pneumocytes (189).

Many viruses enter the cell via endocytosis, and a number of candidate SLiMs relevant for SARS-CoV-2 infection are related to endocytosis (see above). Chlorpromazine, an antipsychotic dopamine D2 antagonist developed in the 1950s, is a potent endocytosis inhibitor (which likely explains its reputation as a “dirty drug” and some of its marked side effects, which can include low white blood cell levels). Like other tricyclic antipsychotics, the drug specifically inhibits the dynamin motor protein that is required to close off the endocytic vesicle at the plasma membrane (195). Anecdotally, it is thought that chlorpromazine (and presumably other tricyclic antipsychotics) might be protecting patients in psychiatric hospitals, and a clinical trial is now planned for COVID-19 (196). The potential use of endocytosis inhibitors such as amiodarone (197) and chlorpromazine in coronavirus infection is further discussed elsewhere (198). Another drug candidate, apilimod, acts later in the endosomal pathway by inhibiting the phosphoinositide kinase PIKfyve and blocks the SARS-CoV-2 entry pathway (38). Apilimod has been shown to strongly inhibit SARS-CoV-2 infection in two additional studies with the half-maximal inhibitory concentration of apilimod observed at 2.5 μM (199). Nafamostat mesylate is active against TMPRSS2 and blocks cell entry (4). As part of the United Kingdom’s ACCORD (Accelerating COVID-19 Research & Development) program, a clinical trial is underway to evaluate bempidainib, a specific inhibitor of the receptor tyrosine kinase AXL in COVID-19 (190). AXL acts as a pleiotropic inhibitor of innate immunity (191) and is also a receptor for Ebola virus (192).

A number of protease inhibitors are now discussed for SARS-CoV-2 treatment. Serine protease inhibitor camostat mesylate is active against TMPRSS2 and blocks cell entry (4). Nafamostat mesylate—originally developed as a trypsin inhibitor (193)—also has been shown to inhibit TMPRSS2. Nafamostat mesylate is an approved anticoagulant in Japan, with clinical testing for COVID-19 infections now being conducted. The spike protein of SARS-CoV-2 contains a furin cleavage sequence (PRRARS|V). Consequently, furin convertase inhibitors are considered as antiviral agents (194). A prime example of such inhibitors is decanoyl-RVKR-CMK, which has been shown to inhibit cleavage of the SARS-CoV-2 spike protein at the S1/S2 site by furin (90). A large drug screen identified four drugs that targeted host cysteine proteases in SARS-CoV-2–infected human cells including VBY-825 (cathepsin B/L), ZLVG CHN2, ONO 5334 (cathepsin K), and MDL-28170 (cathepsin B and calpain I/II), with the latter two inhibiting SARS-CoV-2 replication in human induced pluripotent stem cell (iPSC) pneumocytes (189).

Many viruses enter the cell via endocytosis, and a number of candidate SLiMs relevant for SARS-CoV-2 infection are related to endocytosis (see above). Chlorpromazine, an antipsychotic dopamine
concentration (IC$_{50}$ 0.007 μm) in human lung cells (199) and the half-maximal response (EC$_{50}$ 0.023 μm) in Vero E6 cells (189). Tyrphostin AG 538 (originally described as an inhibitor of insulin-like growth factor 1 receptor) is a preclinical inhibitor of the phosphoinositide kinase PI5P4Kα (gene name PIP4K2A) (199) and showed 55% reduction of SARS-CoV-2 infection in Vero E6 cells (189). PIKfyve generates PtdIns5P, a regulatory phospholipid found on intracellular membrane systems including endosomes (201), whereas PIP4Kα removes PtdIns5P by converting it to PtdIns(4, 5)P$_2$. Thus, apilimod and tyrphostin AG 538 might be targeting aspects of the same regulatory system controlling endosomal fates.

The threonine-specific AP2-associated protein kinase 1 (AAK1) phosphorylates the μ2 subunit of the AP2 complex, promoting clathrin-mediated endocytosis (107). The related cyclin G–associated kinase (GAK) is also a regulator of endocytosis (202). It has been suggested that baricitinib, which is an AAK1 and GAK inhibitor in

| Table 2. Drugs acting on various processes involved in viral entry and infection. |
|---|---|---|---|---|
| **Name of drug** | **Mode of action** | **Clinical status** | **COVID19 ClinicalTrials.gov IDs** | **Other details** | **ChEMBL ID** |
| Inhibitors of viral attachment | | | | | |
| Camostat mesylate | TMPRSS2 inhibition | Approved (Japan) | NCT04355052, NCT04374019, NCT04353284, NCT04321096, NCT04470544 | Shown to be relevant in pancreatic fibrosis | 590799 |
| Nafamostat mesylate | TMPRSS2 inhibition | Approved (Japan) | NCT04473053 | Also inhibits human tryptase. Has strong anticoagulant effect | 3989553 |
| Decanoyl-RVKR-CMK | Furin inhibition | Preclinical | – | Has been shown to inhibit CoV-2 spike cleavage at S1/S2 site by furin. Because the NRP1 receptor binds after cleavage, furin inhibition is potentially more relevant | 3126388 |
| Abituzumab | Integron inhibition (α₅β₅, pan-α₅) | Phase 2 in oncology | – | May also be relevant in fibrosis | 2109621 |
| Cilengitide | Integrin inhibition (α₅β₃, α₅β₅, α₅β₁) | Phase 3 in oncology | – | May also be relevant in fibrosis, sepsis | 429876 |
| Suramin | Entry/early replication inhibitor, mode of action unknown | FDA approved | – | Drug of choice for treating African trypanosomiasis. Not without side effects | 413376 |
| **Endocytosis inhibitors** | | | | | |
| Amiodarone | Inhibits late endosomes | FDA approved | NCT04351763 | Cell culture–based evidence for SARS-CoV inhibition | 633 |
| Chlorpromazine | Inhibits dynamin, thereby blocking clathrin-mediated endocytosis | FDA approved | NCT04354805, NCT04366739 | Antipsychotic drug routinely used as endocytosis inhibitor in cell culture | 823 |
| Apilimod | Inhibits PIKfyve, a regulator of endosomal trafficking | Phase 2 | NCT04446377 | Blocks SARS-CoV-2 infection. Good safety profile from (failed) clinical trials for immune disease application. Clinical trials are registered | 4297643 |
| Imatinib | Abl inhibitor | FDA approved | NCT04422678, NCT04394416, NCT04346147, NCT04357613 | First-line treatment for chronic myeloid leukemia | 941 |
| Teicoplanin | Inhibits cathepsin L (late endosomes and lysosome) | FDA approved | – | Glycopeptide antibiotic | 2367892 |

continued on the next page
addition to being a Janus kinase (JAK) inhibitor, be tested in COVID-19 (203).

The situation with targeting autophagy seems unclear. Autophagy activators might help the cell to consume incoming virus or speed up the establishment of the viral replication complexes and accelerate disease. Autophagy inhibitors might work in later stages of infection to dampen viral production, but this will depend on whether autophagy is active at the time or if the constituent components have been captured and effectively shut down. Several inhibitors/activators have been reported (a selection is listed in Table 2), which can target autophagy and multiple auxiliary signals feeding into the process of autophagy (204–207). One such axis is via the mTORC1 pathway.

### Table 2: Autophagy Modulators

<table>
<thead>
<tr>
<th>Name of drug</th>
<th>Mode of action</th>
<th>Clinical status</th>
<th>COVID19 ClinicalTrials.gov IDs*</th>
<th>Other details</th>
<th>ChEMBL ID†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baricitinib</td>
<td>Inhibits AAK1 and GAK endocytic kinases</td>
<td>FDA approved</td>
<td>NCT04321993, NCT04340232,</td>
<td>Also inhibits JAK kinase</td>
<td>2105759</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NCT04401579, NCT04362943,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NCT04421027, NCT04390464,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NCT04358614, NCT04346147,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NCT04373044, NCT04393051,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NCT04320277, NCT04399798</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BI-853520</td>
<td>FAK inhibitor</td>
<td>Phase 1</td>
<td>–</td>
<td>FAK has been implicated in the Influenza A virus cell entry and replication§</td>
<td>3544961</td>
</tr>
<tr>
<td>Saracatinib</td>
<td>Src and Abl inhibitor</td>
<td>Phase 2 in oncology</td>
<td>–</td>
<td>Now also considered for Alzheimer’s disease</td>
<td>217092</td>
</tr>
<tr>
<td>Tyrophostin A9</td>
<td>PDGF receptor kinase inhibitor (plus other activities)</td>
<td>Preclinical</td>
<td>–</td>
<td>Inhibits actin ring formation</td>
<td>78150</td>
</tr>
</tbody>
</table>

### Autophagy modulators

<table>
<thead>
<tr>
<th>Name of drug</th>
<th>Mode of action</th>
<th>Clinical status</th>
<th>COVID19 ClinicalTrials.gov IDs*</th>
<th>Other details</th>
<th>ChEMBL ID†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metformin</td>
<td>NDUF modulation; mTOR pathway modulation (plus other activities?)</td>
<td>FDA approved</td>
<td>NCT04510194</td>
<td>Approved for type 2 diabetes</td>
<td>1431</td>
</tr>
<tr>
<td>Rapamycin</td>
<td>mTORC1 inhibition</td>
<td>FDA approved</td>
<td>NCT04482712</td>
<td>Used for preventing transplant rejection</td>
<td>1908360 413</td>
</tr>
<tr>
<td>Simvastatin</td>
<td>Autophagy up-regulation via mTOR</td>
<td>FDA approved</td>
<td>NCT04348695, NCT02735707</td>
<td>Treatment for dyslipidemia and atherosclerosis prevention</td>
<td>1064</td>
</tr>
<tr>
<td>Niclosamide</td>
<td>Inhibits SKP2</td>
<td>Niclosamide—FDA approved; VAL—Preclinical</td>
<td>NCT04372082, NCT04436458, NCT0399356</td>
<td>Niclosamide—moderate effect; VAL targets MERS-CoV in cell culture,§ known to inhibit SKP2</td>
<td>1448</td>
</tr>
<tr>
<td>Valinomycin (VAL)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NVP-BEZ235/dactolisib</td>
<td>Autophagy induction</td>
<td>Phase 2</td>
<td>P3K/Akt/mTOR</td>
<td>Inhibits dengue virus replication in tissue culture model,** Reported to have low toxicity in mice††</td>
<td>1879463</td>
</tr>
<tr>
<td>Spautin-1</td>
<td>Autophagy down-regulation by inhibition of USP10 and USP13</td>
<td>Preclinical</td>
<td>–</td>
<td>Inhibits SARS-CoV-2 replication in Calu-3 and Vero E6 cells‡‡</td>
<td>2391504</td>
</tr>
<tr>
<td>Fluoxetine</td>
<td>Affects endolysosomal acidification and cholesterol accumulation in the endosomes</td>
<td>FDA approved</td>
<td>NCT04377308</td>
<td>Used in treatment of major depressive disorder and obsessive compulsive disorder. Inhibits SARS-CoV-2 replication in Calu-3 and Vero E6 cells‡‡</td>
<td>41</td>
</tr>
</tbody>
</table>

DISCUSSION
Examination of viral and cellular proteins that are known (or likely) to be involved in SARS-CoV-2 cell entry has proved a fruitful exercise in identifying multiple candidate SLiMs that might partake in the process. Because of the low sequence information content of SLiMs, it is difficult to get strong statistical support, and therefore, amino acid conservation over long evolutionary time periods is one of the most critical observations in this work. Experimental follow-up is essential before a SLiM can be considered to be functional. A first experimental step to indicate that these motifs might be functional is to test in vitro their ability to bind partner protein domains with an expected low micromolar or high nanomolar affinity, as is typical for SLIM interactions. In the case of PTB/apoPTB motif variants, it is difficult to do this quickly as the binding domain families are large and have not been systematically analyzed for motif preference, despite the known importance of this domain family in endocytosis and vesicle trafficking (136). However, in the accompanying experimental paper, Klische et al. (48) were able to test peptide-domain binding for most of the newly predicted motifs in the ACE2 and integrin tails. Of those that could be quickly tested, several bound with plausible affinities, some were ambiguous, and some did not bind under the conditions of the experiments. Therefore, the ACE2 PBMs, the SPK SH2-binding motif, the AP μ2-binding motif, as well as the integrin β3 phospho-LIR, are now available for their potential roles to be studied in cellular endocytic-autophagosome pathways both under normal conditions and when under pathogenic abuse.

Our observations of SLIM candidates in the viral entry system reveal previously unknown possible interactions mediating viral infection and reflect additional areas of the cell where drug repurposing for HDT might be explored. However, understanding the SARS-CoV-2 entry system has become more complicated and also more confusing. ACE2 is considered the canonical receptor for both SARS-CoV and SARS-CoV-2. Yet, the very rare expression of ACE2 in lungs implies that it cannot be the receptor underlying the severe lung pathology. Integrins are available to play this role and, as we have discussed, are promising but not yet conclusively proven receptors. NRP1 has also been shown to be a SARS-CoV-2 receptor (91) and could also play a role in the lung disease. Like ACE2 and β integrins, we note that this newly identified receptor has a conserved apoPTB motif candidate and also a C-terminal SEA motif that binds the PDZ domain of GIPC1 (212), an endocytosis adapter protein (fig. S8A). GIPC1 can also bind an SDA motif in integrin α (213). Integrins and NRP1 are involved in co-regulated signaling (text S1) and could therefore work together as SARS-CoV-2 receptors. The possibility of additional receptors should not be excluded. Earlier work with SARS-CoV has shown that several plasma membrane lectins can act as co-receptors (32), and in SARS-CoV-2, lectins expressed by innate immune cells bind the spike protein with high affinity and can promote viral entry (214). The cytosolic tails of a number of the lectin receptors are substrates of tyrosine kinases. Although tyrosine kinase inhibitors have frequently been shown to dampen pathogen invasion and disease progression in cell culture, there has been little effort to move these findings into the clinic (24). Because of their widespread use in cancer, the safety profiles of tyrosine kinase inhibitors are well known and we wonder whether this might be a neglected opportunity. COVID-19 might now drive the medical research forward: For example, in a small-scale trial, acalabrutinib, an inhibitor of BTK (Bruton’s tyrosine kinase) that is an activator of macrophages, appears to be effective in dampening excessive inflammation in the lungs of COVID-19 patients (215) and a larger trial is underway (190).

Drugging the cell to cure the pathogen using HD1s is unlikely to fully remove a virus. This would also be undesirable, because the immune system must mount a defense to prevent viral re-infection. Rather, dampening viral load during viral invasion or replication should be the target to give the host defenses time to respond. It is well known that drugs like Tamiflu, which slows influenzaexit and therefore entry into uninfected cells, can only have a strong effect when taken prophylactically or early in infection (216). Depending on the importance of integrins in SARS-CoV-2 lung cell entry, given that it has become clear that ACE2 is barely expressed in the lung, altering virus-cell interactions is a possible role for cilengitide or other molecules that hamper integrin binding. An endocytosis inhibitor might play a similar role and is independent of receptor type. However, for any such inhibitor that passes the blood-brain barrier, effects on mood and other brain operations are an inevitable side effect: Even so, the endocytosis inhibitor chlorpromazine is a widely used drug with a well-known safety profile (217). AAK1 inhibition is also known to affect endocytosis, but there are no specific inhibitors in human trials, leading to the suggestion that the less specific JAK inhibitor baricitinib be tested (203).

Because of the presence of the cell attachment motif RGD in SARS-CoV-2, integrin inhibitors seem worthwhile to explore further. Cilengitide [a cyclic peptide that proved safe in patients but failed to show a survival benefit in glioblastoma (182)] is a relatively selective integrin αβ inhibitor, with low activity on various other integrins. It might be useful in two phases: It could block virus attachment to target cells, and it has also been proposed as a potential treatment in sepsis by preventing the attachment of bacteria to endothelial cells that can be RGD dependent (183). Sepsis was the most frequently observed complication among COVID-19 patients in Wuhan (218). Another potential application of integrin inhibitors, especially integrin αβ, would be lung fibrosis—patients on respirators tend to develop lung fibrosis and show increased αβ levels (219). The antibody abituzumab (DI-17E6) is a pan-α antibody.

with high potency on α6β4, i.e., also active against several α6 integrins, and may consequently be better suited for blocking virus entry or may be suitable for lung fibrosis. It has been clinically tested in several cancer indications (184, 185), and proved safe, but did not achieve a survival benefit. Very recently, the integrin α6β4 inhibitor GS3008348 has been shown to have an effect on lung fibrosis in a mouse therapeutic model (220).

The first large-scale proteomics study expressed 26 tagged SARS-CoV-2 proteins individually in human embryonic kidney (HEK) 293T cells, using affinity purification mass spectrometry to create a viral-human protein-protein interaction map and identified FDA-approved drugs, which were screened as inhibitors of these interactions (209). However, because this screen was not performed in the context of viral infection, it might have failed to reveal interactions essential for viral entry. The drug assays revealed positive results for inhibitors of mRNA translation and modulators of Sigma1/2 receptors and returned negative results for serine protease inhibitors (camostat and nafamostat), ACE inhibitors (captopril and lisinopril), mTOR inhibitors (rapamycin and sapanisertib), and a limited subset of tyrosine kinase inhibitors (midostaurin, ruxolitinib, and pazopanib). Translation and splicing regulators were further found as positive hits in another large-scale proteomics and translome study (221).

Another large drug screen using A549 lung cells identified AXL and AKT kinase inhibitors gilteritinib and ipatasertib and metallocprotease inhibitors primonastat and marimastat as promising drug candidates (222). A phosphoproteomic screen revealed that CK2, p38/MAPK, and cell cycle kinase pathways were regulated by SARS-CoV-2 infection, identifying kinase inhibitors that worked to inhibit viral replication. However, because tyrosine kinases were not strongly represented in the studied dataset (199), future work should help establish the role played by the tyrosine kinase signaling pathway in SARS-CoV-2 infection. While the positive hits in the latter studies provided useful targets for follow-up, the fact that the drug testing was performed in Vero E6 cells implies that negative results should be taken with caution until infection assays reflecting more physiological conditions are tested. A more recent large-scale drug screening found positive results for two tyrosine kinase inhibitors and confirmed strong inhibitory potency in lung iPSCs for two cysteine protease inhibitors and apilimod, with the latter also working in lung explant models (189). The three strongest inhibitors in this study affected viral entry, with apilimod showing potency comparable to that of remdesivir. This suggests that combination therapies using remdesivir together with apilimod or other kinase inhibitors depending on the infection stage might provide promising avenues for fighting COVID-19 infection.

In summary, we have presented evidence at the sequence level for SLiMs in ACE2 and β integrins with the potential to function in viral attachment, entry, and replication for SARS-CoV-2. We have identified several candidate molecular links and testable hypotheses that might help uncover the (still poorly understood) mechanisms of SARS-CoV-2 entry and replication. Because most of these motifs belong to host proteins acting as viral receptors, they are not revealed by virus-centered proteomic interaction assays. That they may well be functional, however, is indicated by sequence conservation, in some cases for hundreds of millions of years. In addition, the motifs are in appropriate cellular contexts to interact with their partner proteins. These putative motifs originally lacked direct experimental evidence, but the accompanying paper from Kliche et al. (48) shows that ACE2 YxxPhi, SH2, and PBM motifs and the integrin β3 phospho-LIR bind to partner domains in vitro. Further experimental follow-up will yield insights into RME for the SARS-CoV-2 virus and, in addition, for the role of ACE2 in the normal cell, where it surely has much more functionality than its role as an ACE. Overall, the collection of candidate motifs in this system suggests that a range of HDTs might be explored including RGD inhibition, tyrosine kinase inhibition, endocytosis inhibition, and autophagy inhibition and/or activation.

**MATERIALS AND METHODS**

**Protein sequences**

Sequences of the spike protein, ACE2, and β integrins, together with homologous sequences, were retrieved from UniProt (222) (release 6 May 2020). The RatG spike protein sequence (Fig. 1) was retrieved from GenBank (224), and sequences of hagfish (*Eptatretus burgeri*) proteins were retrieved from UniParc (UniProt archive). All numberings defining regions in these proteins refer to the canonical sequences. RGD motif and its flanking region (2 amino acids) conservation were checked by using full-length high-quality spike protein sequences retrieved from the GISAID (53, 54) on 9 June 2020.

**Identification of SLiMs**

Candidate and known functional motifs were identified in the analyzed sequences using the ELM web server (23). The tail of ACE2 was also examined for SH2 predictions with ModPeptInt server (116). Structural and context filters were used with their default settings. Motif candidates were manually inspected and were correlated with information from sequence conservation, available structures, and known phosphorylation sites, if applicable.

**Protein alignments, structures, and abundance**

Protein sequences were aligned using Clustal omega (225) using sequences extracted from UniProt and UniParc as inputs. Alignments were visualized using Jalview (226) and Clustal colors. The neighbor-joining tree in Fig. 1B was generated by Jalview with default parameters, and only the sequence regions shown in Fig. 1A were used. Protein structures were retrieved from the PDB (227) and were visualized using UCSF Chimera (228). Expression levels for various human proteins were taken from the Human Protein Atlas (81).

**PTM sites**

PTM sites in ACE2 and integrin β1, β3, and β6 tails were taken from PhosphoSitePlus (v6.5.9.3) (143). Visualization of PTM data was done using the images generated by the PhosphoSitePlus web server using both HTP and LTP phosphorylation data.

**SUPPLEMENTARY MATERIALS**

stke.sciencemag.org/cgi/content/full/14/665/eabd0334/DC1

Text S1. Extended discussion on the potential of NRPs in SARS-CoV-2 cell entry. Fig. S1. The structural feasibility of simultaneous integrin and ACE2 binding by SARS-CoV-2 spike protein trimers. Fig. S2. Structural indication of a functional spike protein RBD/integrin αvβ3 interaction. Fig. S3. Alignment of human ACE2 transmembrane helix and C-terminal intracellular tail with homologous sequences of representative vertebrate collectins from UniProt reference proteomes. Fig. S4. Alignment of homologous sequences of integrin β3 transmembrane helices and intracellular tails. Fig. S5. Alignment of homologous sequences of integrin β6 transmembrane helices and intracellular tails.

REFERENCES AND NOTES


B. Shen, K. Vardy, P. Hughes, A. Tasdogan, Z. Zhao, R. Yue, G. M. Crane, S. I. Morrison, Integrin alpha1 is an osteolectin receptor and is required for the maintenance of adult skeletal bone mass. eLife 8, e42274 (2019).


A transmembrane serine protease is linked to X.


P. sodium-coupled monocarboxylate transporter SMCT1 (SLC5A8) and E. membrane fibronectin receptor with 444–448 (1998). rearrangements induced by the alphaIIbbeta3-integrin signaling. promoting ROCK activation in 3-kinase enhancer-activating Akt, preventing its apoptotic cleavage and integrin activation to (2014). G. m-band protein skelemin and A.  integrin cytoplasmic tail are required for polarize focal adhesion assembly.


Discovery of randomisation platform study to host cell furin proprotein convertases as high-throughput assay for SARS-CoV-2 infection.


Phosphatidylinositol 5-phosphate 4-kinase with potential of simvastatin alleviates airway inflammation and modulating through up-regulation of autophagy in mouse models of asthma. Respir Res. 2015, 16 (1), 53–64.


M. Roebuck, D. S. Cummings, Simvastatin alleviates airway inflammation and modulating through up-regulation of autophagy in mouse models of asthma. Respir Res. 2015, 16 (1), 53–64.

G. M. 14 May 2011. http://stke.sciencemag.org/ Downloaded from
Acknowledgments: We thank J. Klíček, H. Kuss, M. Ali, and Y. Iverson from the Uppsala University for testing several motif candidates and sharing their peptide-domain binding data in advance of publication. We thank S. Michael for helping to prepare the ELM entry LIG_NRPCendR_1. We are grateful to G. Jenkins and the Nottingham Covid Research Group for making integrin and ACE2 data publicly available on laboratory website (https://www.nottinghamcrg.info/results).

Funding: B.M. has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 675341 (PDZNet). E.M.-P. is a PhD student of CONICET, Argentina. R.A. is supported by BMBF-funded Heidelberg Center for Human Bioinformatics (HD-Hub) within the German Network for Bioinformatics Infrastructure (de.NBI #031AS5378) and ELIXIR Germany. L.B.C. is a National Research Council Investigator (CONICET, Argentina). The work was supported by Agencia Nacional de Promoción Científica y Tecnológica (PICT 2017-2192) grant to L.B.C. This paper is part of a project that has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 778247 (IPDPlus) to L.B.C., D.C.S., and T.J.G. IPDPlus also funded E.M.-P.’s placement at EMBL.

Author contributions: All authors worked on the manuscript as it evolved and contributed to the regular discussions by teleconference. B.M. (ORCID: 0000-0003-0919-4449) identified and analyzed integrin-related SLMs, analyzed sequence and structural data, prepared the corresponding alignments and structural figures, and wrote the sections related to integrins. H.S.-S. (ORCID: 0000-0003-4744-4787) contributed to the introduction, the NPY and SH2 motif sections, and the discussion on tyrosine kinase inhibitors. J.-A.-V. (ORCID: 0000-0003-0752-1151) prepared sequence alignments and figures. J.C. (ORCID: 0000-0003-1047-4157) analyzed the ACE2 PDB and wrote the related results section. E.M.-P. (ORCID: 0000-0003-4411-976X) prepared the figure of phosphorylation of ACE2 tail (literature searches for phosphorylation of ACE2 tail). R.A. (ORCID: 0000-0002-7212-0234) undertook literature searches, maintained the reference database, and coordinated with B.M. to meet submission requirements. D.C.S. (ORCID: 0000-0003-4015-2474) worked on the RGD and integrin aspects. M.K. (ORCID: 0000-0002-3004-2151) worked on defining regular expression for autophagy (LIR autophagy), wrote LIR autophagy motif section, contributed toward developing drug table, and wrote parts of the introduction, therapeutics, and discussion section. F.R. (ORCID: 0000-0002-4604-9251) provided cleftidge aspects, including development history, and assessments of other drugs, in the therapeutics section, and discussion. L.B.C. (ORCID: 0000-0003-0192-9906) worked on the identification and definition of ACE2 tail motifs, the assessment of the switching mechanisms acting on the ACE2 tail, and the assessment of drugs targeting viral entry; participated in figure design; and wrote several sections of the introduction, results, and the discussion. T.J.G. (ORCID: 0000-0003-0657-5166) undertook SLIM searches and evaluations, prepared sequence alignments, and wrote the YxxPhI section and parts of the results, introduction, and discussion.

Competing interests: F.R. has worked on integrin inhibitors in Merck KGaA, Darmstadt, Germany. All other authors declare that they have no competing interests.

Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper or the Supplementary Materials.

Submitted 28 May 2020
Accepted 10 December 2020
Published 12 January 2021
10.1126/scisignal.abd0334

Short linear motif candidates in the cell entry system used by SARS-CoV-2 and their potential therapeutic implications

Bálint Mézáros, Hugo Sámano-Sánchez, Jesús Alvarado-Valverde, Jelena Calyseva, Elizabeth Martínez-Pérez, Renato Alves, Denis C. Shields, Manjeet Kumar, Friedrich Rippmann, Lucía B. Chemes and Toby J. Gibson

DOI: 10.1126/scisignal.abd0334

SARS-CoV-2: From entry to autophagy?

SARS-CoV-2, the virus that causes COVID-19, enters cells through endocytosis upon binding to the cell surface receptor ACE2 and potentially others, including integrins. Using bioinformatic analysis, Mészáros et al. predicted the presence of short amino acid sequences, called short linear motifs (SLiMs), in the cytoplasmic tails of ACE2 and various integrins that may engage the endocytic and autophagic machinery. Using affinity binding assays, Kliche et al. not only confirmed that many of these predicted SLiMs interacted with target peptides in various components of the endocytosis and autophagy machinery but also found that these interactions were regulated by the phosphorylation of SLiM-adjacent amino acids. Together, these findings have identified a potential link between autophagy and integrin signaling and could lead to new ways to prevent viral infection.