

Journal Pre-proof

Deamidation drives molecular aging of the SARS-CoV-2 spike protein receptor-binding motif

Ramiro Lorenzo, Lucas A. Defelipe, Lucio Aliperti, Stephan Niebling, Tânia F. Custódio, Christian Löw, Jennifer J. Schwarz, Kim Remans, Patricio O. Craig, Lisandro H. Otero, Sebastián Klinke, María García-Alai, Ignacio E. Sánchez, Leonardo G. Alonso

PII: S0021-9258(21)00977-7

DOI: <https://doi.org/10.1016/j.jbc.2021.101175>

Reference: JBC 101175

To appear in: *Journal of Biological Chemistry*

Received Date: 8 July 2021

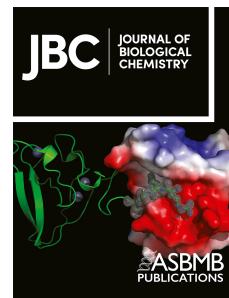
Revised Date: 3 September 2021

Accepted Date: 3 September 2021

Please cite this article as: Lorenzo R, Defelipe LA, Aliperti L, Niebling S, Custódio TF, Löw C, Schwarz JJ, Remans K, Craig PO, Otero LH, Klinke S, García-Alai M, Sánchez IE, Alonso LG, Deamidation drives molecular aging of the SARS-CoV-2 spike protein receptor-binding motif, *Journal of Biological Chemistry* (2021), doi: <https://doi.org/10.1016/j.jbc.2021.101175>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology.



Deamidation drives molecular aging of the SARS-CoV-2 spike protein receptor-binding motif

Ramiro Lorenzo^{1,¶}, Lucas A. Defelipe^{2,¶}, Lucio Aliperti³, Stephan Niebling^{2,4}, Tânia F. Custódio^{2,4}, Christian Löw^{2,4}, Jennifer J. Schwarz⁵, Kim Remans⁵, Patricio O. Craig^{6,7}, Lisandro H. Otero⁸, Sebastián Klinke⁸, María García-Alai^{2,4}, Ignacio E. Sánchez³ and Leonardo G. Alonso^{9,*}

¹ Centro de Investigación Veterinaria de Tandil (CIVETAN), CONICET-CICPBA-UNCPBA, Facultad de Ciencias Veterinarias, Universidad Nacional del Centro (FCV-UNCPBA), Tandil, Argentina.

² European Molecular Biology Laboratory, Hamburg Unit, Hamburg, Germany.

³ Universidad de Buenos Aires, Consejo Nacional de Investigaciones Científicas y Técnicas, Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN), Facultad de Ciencias Exactas y Naturales, Laboratorio de Fisiología de Proteínas, Buenos Aires, Argentina.

⁴ Centre for Structural Systems Biology, Hamburg, Germany.

⁵ European Molecular Biology Laboratory, Heidelberg, Germany.

⁶ Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales. Departamento de Química Biológica. Buenos Aires, Argentina.

⁷ CONICET-Universidad de Buenos Aires. Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN). Buenos Aires, Argentina.

⁸ Fundación Instituto Leloir, IIBBA–CONICET, and Plataforma Argentina de Biología Estructural y Metabólica PLABEM, Buenos Aires, Argentina.

⁹ Instituto de Nanobiotecnología (NANOBIOTEC), UBA-CONICET-Universidad de Buenos Aires, Buenos Aires, Argentina.

¶ These authors contributed equally to this work.

* To whom correspondence may be addressed. E-mail: lonso@docente.fyfb.uba.ar

Running title:

Deamidation drives aging of SARS-CoV-2 spike protein

Keywords:

SARS-CoV-2, receptor-binding, spike, molecular aging, protein deamidation, RNA virus, protein evolution, protein-protein interaction, receptor structure-function.

Abstract

The spike protein is the main protein component of the SARS-CoV-2 virion surface. The spike receptor-binding motif mediates recognition of the human angiotensin-converting enzyme 2 (hACE2) receptor, a critical step in infection, and is the preferential target for spike-neutralizing antibodies. Post-translational modifications of the spike receptor-binding motif have been shown to modulate viral infectivity and host immune response, but these modifications are still being explored. Here we studied asparagine deamidation of the spike protein, a spontaneous event that leads to the appearance of aspartic and isoaspartic residues, which affect both the protein backbone and its charge. We used computational prediction and biochemical experiments to identify five deamidation hotspots in the SARS-CoV-2 spike protein. Asparagine residues 481 and 501 in the receptor-binding motif deamidate with a half-life of 16.5 and 123 days at 37°C, respectively. Deamidation is significantly slowed at 4°C, indicating a strong dependence of spike protein molecular aging on environmental conditions. Deamidation of the spike receptor-binding motif decreases the equilibrium constant for binding to the hACE2 receptor more than 3.5-fold, yet its high conservation pattern suggests some positive effect on viral fitness. We propose a model for deamidation of the full SARS-CoV-2 virion illustrating how deamidation of the spike receptor-binding motif could lead to the accumulation on the virion surface of a non-negligible chemically diverse spike population in a timescale of days. Our findings provide a potential mechanism for molecular aging of the spike protein with significant consequences for understanding virus infectivity and vaccine development.

Introduction

In December 2019, a viral pneumonia outbreak was reported in Wuhan, China (1). This outbreak quickly turned into a pandemic disease (COVID-19) of international concerns (2), and the novel pathogen causative of a Severe Acute Respiratory Syndrome (SARS) was soon

identified as SARS-CoV-2, a new member of the *Betacoronaviruses* genera. SARS-CoV-2 and SARS-CoV, the agent responsible for the 2002-2003 pneumonia outbreak, are closely related to the bat coronaviruses from which they likely originated and passed to an intermediate species that ultimately infected humans (3, 4). The host specificity and infectivity of SARS-CoV-2 and SARS-CoV rely on the spike protein (S). Through its receptor-binding domain (RBD, residues 319 to ~515), S recognizes the human Angiotensin-converting enzyme 2 (hACE2) with nanomolar affinity, triggering events that culminate with the fusion of the cellular and viral membranes (5). In SARS-CoV-2, the S protein is synthesized as a 1273-residue heavily glycosylated polypeptide that is cleaved by the host furin protease between the S1 (1-685) and S2 (686-1273) subunits (6). On the surface of native viruses, the S protein is mainly observed as a metastable trimer in the prefusion conformation. The RBD of each S protomer can switch between a receptor-accessible conformation known as the "up-state" and a receptor-inaccessible and buried conformation that packs against the N-terminal domain (NTD) of the neighboring protomer called the "down-state" (7). Two regions can be identified in the RBD, a conserved core and a more variable region termed as the receptor-binding motif (RBM, residues 438-506). The latter region contains residues that establish direct contact with hACE2, determining S protein affinity and specificity (8).

Inter-species spillover is often observed in the coronavirus family members, a phenomenon that mainly originates from amino acid mutations in the RBD that enables S to bind ACE proteins from two different host species (9, 10). Beyond S crucial role in restricting viral host infectivity, the protein is the target of potent neutralizing antibodies (11–13) with therapeutic use and the main antigenic component of vaccines (14, 15). It is of particular interest to understand how mutations and post-translational modifications (PTM) in RBD affect viral infectivity, generate antigenic escape variants or restrict the humoral and cellular immunity.

Asparagine (Asn) deamidation is a spontaneous and irreversible frequently observed PTM (16, 17). As a result of the substitution in the Asn side-chain of the carboxamide nitrogen atom by

a hydroxyl group, a mixture of aspartic and isoaspartic acid (a beta amino acid) is generated (18), introducing a negative charge and a rearrangement of the protein backbone in the latter case. The deamidation rate, which heavily depends on the primary sequence and local structure, can be estimated using bioinformatic tools that rely on different approaches such as structural constraints, machine-learning or primary sequence and disorder predictors (19–21). The fastest deamidation rate (half-time ~ 1.2 days)

is often observed for the asparagine-glycine (NG) dipeptide in loosely structured regions, which are generally regarded as deamidation "hotspots" for which most of the methods provide accurate prediction (22). In this regard, although asparagine deamidation is a relatively slow process (half-times of hours to days), it can significantly occur within the time frame in which some critical viral processes develop. For instance the SARS-CoV-2 virus can remain infective for a couple of days after incubation at 37 °C (23).

The S protein contains multiple deamidation sites, but only some of them bear biological relevance. To identify these relevant sites, we applied a prioritization process based on four key points: (i) *In-silico* identification of hotspots and deamidation rates using the NGOME-LITE algorithm (24), (ii) conservation of the sites among *Betacoronaviruses*, (iii) location of the sites at the ACE2 binding surface, and (iv) experimental determination of the deamidation rates at specific hotspots by mass spectrometry. A direct consequence of deamidation events in Asn residues located at the RBM is the modification of the electrostatic potential of the receptor-binding surface due to the introduction of a negative charge from the aspartic or isoaspartic carboxylate. Consequently, as the RBD is the preferential region from the S protein for neutralizing antibodies (11–13), deamidation could also affect the efficiency of the humoral immune response. With these data we generated a model that computes the proportion of deamidated protomers in a single viral particle and predicts how these species evolve. Our findings shed light on deamidation, an aging mechanism that operates on the RBM, a critical region that determines virus infectivity and may profoundly impact on vaccine development.

Results

Identification of deamidation hotspots in S proteins

We initially estimate the individual deamidation (Fig 1A) half-time ($t_{1/2}$) using the NGOME-LITE (24) algorithm, for all Asn residues with no glycosylation probability for a set of *Betacoronavirus* S proteins (Fig 1B). We restricted our analysis to a group of S proteins from *Betacoronaviruses* with a demonstrated affinity for the hACE2 receptor, which includes SARS-CoV-2 (Wuhan), SARS-CoV (Urbani), SARS-CoV (GZ0402), the closely related Bat SARS-like CoVs RaTG13 (25) and WIV1 (26) (Table S1).

A similar deamidation profile is observed for the five S proteins (Fig 1B), characterized by a major group of Asn residues in a slow-deamidation regime with estimated $t_{1/2}$ greater than 100 days and a restricted group of residues in a fast-deamidation regime, with estimated $t_{1/2}$ ranging between 20 to 35 days (Fig 1B and Table S2). All Asn residues in the fast-deamidation regime (hereafter referred to as deamidation hotspots) are part of NG dipeptides (Fig S1), for which a deamidation $t_{1/2}$ of 1.2 days is expected in the absence of structural protection (*i.e.* in a short non-structured peptide) (18, 19) indicating that the S protein fold protects these Asn residues from deamidation. Slow regime residues virtually do not deaminate during the viral-life cycle and are devoid of any functional consequence.

Except for the single deamidation hotspot observed at position 74 in the Bat-CoV RaTG13 (in SARS-CoV-2 Asn residue 74 is glycosylated (27)), deamidation hotspots are restricted to a group of residues clustered between residues 481 and 501 in the RBM, including residue 493 (only observed in the GZ0402 SARS-CoV), which shows partial conservation among analyzed CoVs, and a group of strictly conserved residues at positions 544, 856 and 907 (Fig 1C and Fig S1), observed in all the selected S sequences. Hotspots within the RBM are, altogether, termed as the *RBM deamidation cluster*. Overall, the deamidation profile shows that hotspots are not evenly distributed along with these S protein sequences (Figs 1B and 1C).

Beyond sequence context, the protein fold, backbone flexibility and surface accessibility

correlate with the experimental Asn deamidation rates (28, 29). The hydrolysis of the cyclic intermediate is required to complete the deamidation reaction (Fig 1A) for which solvent molecules must reach the reaction center. Solvent accessibility can be assessed by calculating the Relative Accessible Surface Area (RASA) of the deamidation-prone residue to a small probe radius of 1.4 Å (a water molecule). In addition, NGOME-LITE cannot sense the effect of the tertiary and quaternary structure on the deamidation rate, factors that become of critical importance for hotspots centered at the inter-domain or inter-protomeric interfaces of the S trimer. Higher-order structure factors affecting deamidation rates can be identified by evaluating residue accessibility to a 3.0 Å radius probe.

To discriminate between the predicted deamidation-prone residues that are solvent accessible and those present at the inter-domain surface we calculated the side-chain RASA for all deamidation hotspots in the SARS-CoV-2 S structure. We performed our analysis using a structure of the furin-cleaved S protein in the pre-fusion conformation, which contains two RBDs in the down-state and one in the up-state (PDB 6zgg (25) using two probes with radii of 1.4 and 3.0 Å (30), Table S3).

All five deamidation hotspots are fully or partially accessible to the 1.4 Å radius probe (Table S3), albeit to a different extent depending on the RBD conformation, indicating that they are accessible to water molecules, a requirement for hydrolysis. On the other hand, only hotspots 481 and 501 are accessible to the 3.0 Å radius probe in both the up and down RBD conformations (RASA of 74.2% and 25.8% in the up conformation for Asn 481 and 501, respectively), (Fig 1D) while Asn 544 is virtually buried in the RBD in the down-state and partially exposed (RASA of 25.6%) in the up-state (Fig 1E).

Additionally, Asn 907 located in the heavily glycosylated stalk of S is partially accessible to the 1.4 Å radius probe through an internal channel formed by the three protomers, whereas Asn 856 is packed between the contact interface of two protomers (Fig 1C). Both positions are fully inaccessible to the 3.0 Å radius probe, irrespective of the RBD conformation.

Assessment of deamidation $t_{1/2}$ for the 481, 501 and 544 hotspots in mild conditions

The deamidation profile of SARS-CoV-2 S obtained with NGOME-LITE shows the presence of conserved deamidation hotspots at relevant protein-protein recognition interfaces. However, the fact that these sequence stretches are part of a large multidomain protein, which is not detected by the NGOME-LITE algorithm, reduces the prediction accuracy of this bioinformatic tool.

First experimental observation of the occurrence of deamidation at the predicted RBD hotspots was reported previously by the Wells lab (27). To monitor N-glycan occupancy in the full-length S protein, they observed 18.9% of ^{18}O -Asp conversion at Asn 501, 4.8% at Asn 481 and 7.8% at Asn 544 attributable to deamidation, since these hotspots lack canonical glycosylation sequons (N-X-S/T) and were shown to have less than 5% of glycan occupancy. The amount of ^{18}O -Asp conversion correlates well with the deamidation $t_{1/2}$ predicted by NGOME-LITE for all five deamidation hotspots (Table S4).

As the deamidation kinetics is critical to evaluate the potential effect of the aspartic/isoaspartic formation in the S function, we experimentally assessed the deamidation $t_{1/2}$ for the hotspots 481, 501 and 544 in a recombinantly expressed extended version of the SARS-CoV-2 RBD construct encompassing residues 319 to 566 (31) at 4 °C and 37 °C.

It should be noted that residue 544 is not part of the RBD but instead it belongs to the structural conserved subdomain 1 (SD1) that encompasses residues ~516 to ~591 not fully present in our construct and the SD1 may be devoid of its native structure. Consequently, the $t_{1/2}$ for the 544 hotspot is not further considered for functional conclusions. However, it is shown as a corroboration that the technique is adequate for determining deamidation half-times in the expected time lapse. In our experimental setup, the recombinant RBD protein was incubated at pH 7.4 over several days at two different temperatures and the presence of deamidated species was identified and quantified by mass spectrometry (32) (Fig S2 and Table S5). Fig 2A shows the time-decay of the unmodified peptides (Asn containing peptides) bearing the deamidation hotspot 481, 501 and 544. Full data are available via

ProteomeXchange with identifiers PXD028071 (peptide: IYQAGSTPCNGVE) Dataset S1 and PXD027873 (peptides: CVNFNENGLTGTGVLTE and GFNCYFPLQSYGFQPTNGVGYQPYPYR) Dataset S2. Initially (time = 0), the unmodified peptides covering the 481, 501 and 544 hotspots account for 95.6%, 100.0% and 84.0 % of the detected peptides, respectively (Fig 2A and Table S4); these values dropped to 34.9%, 87.9% and 17.4 % for the 481, 501 and 544 hotspots, after 20 days of incubation at 37 °C.

Deamidation half-times of 16.5 ± 3.7 , 123 ± 23 , and 7.9 ± 1.2 days were obtained at 37 °C for the 481, 501 and 544 hotspots in the SARS CoV-2 RBD protein. Under our experimental conditions, the RBD hotspot 501 remains stable against the deamidation process despite being identified *a priori* by the NGOME-LITE as the fastest deamidation hotspot. On the contrary, Asn 544 has the highest deamidation rate, with a value close to the expected deamidation rate of a NG sequence in an unstructured peptide model (1.2 days) supporting that this residue is placed in an unstructured region. The disagreement between experimental determined deamidation rates and computational predictions for hotspots that share the common NG sequence highlights that protein structure and associated dynamics in S can fine-tune the deamidation rate.

As expected, the deamidation reaction was slowed down considerably by decreasing the temperature to 4 °C (Table S4). The deamidation half-time for Asn 481 at 37 °C compared to 4 °C is more than 20-fold higher, highlighting that temperature can critically affect the identity of the RBM.

Deamidated species for hotspots 481 and 544 elute at two different retention times (RT) in a reverse-phase (RP) chromatography experiment (Fig S2), in agreement with the reaction mechanism that proposes the generation of a mixture of aspartic or isoaspartic bearing peptides. In RP chromatography, deamidated species containing isoaspartic residues are reported to elute faster than aspartic containing peptides (18) and, based on this chromatographic behavior, the deamidation reaction (20 days at 37 °C) at hotspots 481 and 544 shows an abundance of 46.6% and 68.5% of isoaspartic containing species, respectively (Table S4). On the contrary, only one peak was observed for the deamidated species at Asn

501.

We then evaluated how deamidation affects the affinity of RBD for the ectodomain of hACE2. To this end we performed a bio-layer interferometry (BLI) assay using an aged (20 days at 37 °C) and heterogeneously deamidated RBD sample with ~65%, ~12% and ~83% of hotspots 481, 501 and 544 in its deamidated form. We observed that the unaged RBD binds hACE2 with an affinity dissociation constant (K_d) of 23.4 ± 0.8 nM whereas the aged RBD has a reduced affinity of only 81.4 ± 0.8 nM (Fig 2B) showing that deamidation is detrimental for hACE2 binding and might impact virus infectivity.

To put this result into context, we must consider that the aged protein is a heterogeneous mixture of different chemical species, where, in a single RBD molecule, different combinations of non-deamidated and deamidated hotspot in their aspartic or isoaspartic forms coexist. Assessing how each chemical species (each individual hotspot in its aspartic or isoaspartic form) contributes to hACE2 binding, is a non-trivial experimental task. The major drawback is the fact that deamination mainly introduces an isoaspartic residue into the hotspot, which produces, beyond the charge effect, a significant rearrangement of the protein backbone, making it difficult to emulate it faithfully by a direct aspartic mutation. Likewise, and although deamidation is the main post-translational modification observed in the aged protein, other less frequent PTMs (such as methionine oxidation) could contribute, in some proportion, to the decrease in the affinity for hACE2.

On the other hand, deep mutational scanning experiment performed by Bloom lab (33) describes the effect of individual aspartic mutations at position 481 and 501. This experiment shows a marginal perturbation in the apparent dissociation constants ($K_{D,app}$) of the RBD-hACE2 complex for the N481D mutation ($\Delta \log_{10} K_{D,app} -0.07 \pm 0.01$; $K_{D,app} \text{ Mut}/K_{D,app} \text{ Wild type} \sim 1.2$ fold), but a significant decrease in the interaction strength ($\Delta \log_{10} K_{D,app} -2.42 \pm 0.03$; $K_{D,app} \text{ Mut}/K_{D,app} \text{ Wild type} \sim 260$ fold) for the N501D mutation (33). We observed a 3.5 fold decrease in the K_d for an aged RBD sample containing ~65% of deamidated hotspot 481, harboring mostly an isoaspartic residue (~47%, Table S4), and ~12% of deamidated hotspot

501 harboring only an aspartic residue (Table S4). Taking these results together it is likely that deamidation at each RBM hotspot additively affect hACE2 binding, with an overwhelming importance of 501 position.

Topological constraints drive localization of conserved deamidation hotspots at the RBM

Due to the critical role of the RBM residues in determining hACE2 affinity and host-specificity (34, 35), we further evaluated the conservation pattern of hotspots at the RBM deamidation cluster by extending our analysis to a group of 38 *Betacoronaviruses* representative of the subgenus *Sarbecoviruses* (36, 37) which, in addition to SARS-CoV and SARS-CoV-2, includes numerous bat and pangolin viruses (Fig 3A, Dataset S3, File S1 and Table S6). We evaluate the correlation between Asn conservation among the selected *Sarbecovirus* versus the estimated deamidation half-time of Asn residues in the SARS CoV-2 S protein and the results are highlighted in Fig 3A.

Overall, the RBM deamidation cluster shows a variable conservation pattern (Figs 3A and 3B) compared to hotspots 544, 856 and 907 which are identified as red dots in the left upper region of Fig 3A indicating that these Asn are both conserved and deamidation-prone residues. On the other hand, hotspots within the RBM deamidation cluster are differentiated based on their opposite conservation pattern. While the hotspot 493 occurs only once in the SARS-CoV GZ0402, and hotspot 501, identified as a green dot in the Fig 3A, shows low conservation (observed in 5 out of 38 sequences) and hotspot 481 (blue dot in Fig 3A) is highly conserved and present in 30 out of 38 sequences (~80% conservation).

Moreover, the RBMs of the *Sarbecovirus* members can be differentiated in two groups according to the presence of a *receptor binding ridge* (RBR), which involves residues ~470 to 488 and includes a disulfide bridge between Cys 480 and 488 (37), highlighted in purple in Fig 3B. Altogether, deamidation hotspots in the RBMs containing a RBR are observed within an 11 amino acid stretch and are restricted to any of the following three positions: 481, 493 and 501 (Fig 3B). The Bat CoVRs/YN2018A, included as a representative *Sarbecovirus* that lacks a RBR and does not bind hACE2, shows a predicted

deamidation hotspot at position 487 (Table S7) that cannot be easily aligned with the RBR bearing sequences (Fig 3B). However, the superposition of the RBDs from SARS-CoV-2 (pdb: 6M0J) and Bat CoVRs/YN2018A obtained by homology modeling shows that the 481 and 487 hotspots are located in a similar position in the three-dimensional structure of the domain enabling sequence alignment between both groups. This suggests that topology drives conservation of deamidation-prone Asn residues at this specific position (Fig 3D).

The overall folds of SARS-CoV and SARS-CoV-2 RBMs are similar with 14 residues (35) observed in equivalent positions in both structures located at less than 5Å from hACE2 binding residues (cut-off for considering a direct contact) including the 493 and 501 deamidation hotspots (Fig 3B). Asn 493 and 501 form a direct interaction with hACE2 residues K31 and K353, respectively, which are considered critical for S binding. Mutational studies (34, 35, 38) have shown the relevance of these two positions in determining hACE2 affinity. In particular, the presence of aspartic acid or asparagine at position 501 in the spike protein determines the possibility of infecting cells through interaction with hACE2. The RaTG13 CoV and SARS CoV-2 share 89.2% of identity at the RBD, however the RaTG13 spike protein does not enable an efficient hACE2-mediated infection in a pseudovirus assay (39, 40). Of the six different positions between RaTG13 and SARS CoV-2 that most contribute to the hACE2 binding (Y449, F486, Q493, Q498, N501, and Y505 in SARS CoV-2), the most important is the aspartic acid at position 501 in the RaTG13 RBD. A single D501N mutation in the RaTG13 increases nine fold the affinity for hACE2 (39) and restores infectivity in a cell-cell fusion assay (40) pinpointing the biological importance of this deamidation hotspot.

On the other hand, the deamidation hotspot at position 481 is fully exposed in the RBD, positioned in the external wall of the RBR (Figs 3C and D) and does not directly participate in hACE2 binding.

Taken together the sequence and topological conservation patterns and the measured deamidation half-times we conclude that deamidation hotspot at the RBM are subjected to different selection forces. The hotspots 493 and 501 which significantly contribute to hACE2

binding are not highly conserved among *Sarbecoviruses* and are not significantly deamidated under our experimental conditions. The occurrence of a deamidation event at these hotspots would be highly detrimental for receptor binding. Furthermore, and of note, the N501Y mutation that eliminates the deamidation hotspot is a hallmark of the emerging SARS-CoV-2 alpha (VOC 202012/01, B.1.1.7) and betha (501Y.V2, B.1.351) variants that are rapidly spreading in the UK and South Africa (41). Both variants are likely to have arisen independently and are associated with increased transmissibility.

A different mechanism might have shaped the presence of a deamidation hotspot at position 481 or equivalent positions that are topologically conserved beyond the absence of an entire structural element (the RBR) and that readily deamidated under physiologically relevant conditions. Although deamidation is often conceived as detrimental for protein function or stability, our findings support the striking possibility that deamidation mediates an aging-dependent gain-of-function in S protein.

Kinetic model for S protein deamidation in the SARS-CoV-2 virion

As spontaneous Asn deamidation is, compared to other modifications that regulate protein function, a slow process, we focused on finding to what extent deamidation of the S protein RBM takes place in the context of the SARS-CoV2 life cycle. We have contextualized the results obtained for Asn 481 and 501 by numerical simulation and graphical representation of the SARS-CoV2 virion. We used the Gillespie algorithm as implemented in COPASI (42) to simulate the stepwise irreversible transition of the 33 S protein trimers in a virion (43, 44) from the intact state to the fully deamidated state. Our model considers deamidation of only Asn 481 and 501, with independent experimental half-time at 37 °C (Fig 2). The six deamidation sites in each S trimer lead to 2^6 possible deamidation states, which can be grouped into 20 species using symmetry considerations (Fig S3).

We performed 1000 stochastic simulations and reported the average and standard deviation of the results for each time point. Fig S4 shows the full results, while Table S8 highlights the results at several time points of interest. The evolution of the different deamidated species of Spike

trimers present in a virion at 37 °C is shown in Fig 4A, grouped by the total number of deamidated sites in the trimer for clarity. The intact S trimers decay with a half time of approximately four days. This value is heavily affected by temperature and the intact S trimer $t_{1/2}$ decay at 4 °C is delayed to 110 days. Due to the multimeric nature of both the virion and the Spike protein, trimers with at least one deamidated hotspot increase early: after only one day, close to 4 S trimers would host a deamidated hotspot (mainly at Asn 481). After two days, close to 7 of the S trimers have deamidated at one site and close to an additional trimer has deamidated at two sites (Table S8). Fig 4B shows a visualization of the deamidation state of the SARS-CoV2 virion after 48 h (45). Only 17 S protein trimers are shown for clarity (45), while the population of each species is proportional to that in Table S8. Our model supports the existence of a non-negligible population of intra-virion S protomers bearing a deamidated hotspot 481 that accumulate within 24 to 48 h, a time-lapse compatible with the viral life-cycle (i.e the period in which SARS CoV-2 remains infective when incubated at 37 °C, (23)). Of note, the molecular mechanism that enables the emergence of intra-virion deamidated S proteins is conserved among *Sarbecovirus* and located at topologically equivalent positions in the different RBDs. Additionally, it is important to note that deamidation may have significantly altered the surface properties of the virion only two days after synthesis and this may have consequences for immune evasion.

On the other hand, the intact S trimers significantly decrease to four after two weeks, while the full deamidation of the virion takes about 1000 days. This timescale is far slower than that of SARS CoV-2 inactivation in solution at 37 °C, (23), indicating that deamidation of RBM hotspots might not be the main molecular mechanism for viral inactivation.

Discussion

Asn deamidation is a PTM that spontaneously and pervasively occurs in proteins containing the dipeptide sequence NG. The quantitative assessment of the time-dependent accumulation of deamidated species at specific positions in large and heavily post-

translationally modified proteins is a difficult experimental task, and consequently, its potential functional role has been neglected. Deamidation is often conceived, by default, as a degradative reaction that is detrimental for protein function and stability. Exceptionally, deamidation gains importance in biologics and vaccine development when deamidating-prone Asn residues are located, by chance, in protein-protein binding interfaces or in antigenically relevant epitopes that critically affect protein performance. (46–48). With the exception of a handful of well-documented cases (49, 50), the participation of Asn deamidation as a mechanism that ultimately leads to a time-dependent protein gain-of-function is scarce. Here we present biochemical and bioinformatics evidence that (i) deamidation hotspots are a conserved trait in the RBM of *Sarbecovirus*, (ii) the measured deamidation half-times for individual Asn residues enable the accumulation of intra-virion deamidated species during the virus life-cycle, and (iii) conservation is driven by topological constraints. Taken together, these results suggest a yet unidentified functional role for conserved deamidation hotspots.

The deamidation profile of the S proteins from a selected group of *Betacoronaviruses* shows a discrete number of predicted deamidation hotspots that are not randomly distributed over the entire length of the S ectodomain. Instead, deamidation hotspots are observed either fully conserved or clustered within the RBM. The RBM cluster is of particular importance due to its role in hACE2 binding and includes hotspots at positions 481, 493 and 501 that are partially conserved among *Sarbecovirus*.

We experimentally observed that Asn 481 deamidates with a half-time of 16.5 days at 37°C, a reaction that is highly temperature-sensitive. An aged RBD sample in which 65% of the 481 hotspot is observed in its deamidated form showed a 3.5-fold decrease of the K_d for hACE2. However, the idea that deamidated residues are always detrimental for protein fitness collides with the observation that hotspot 481 is highly conserved among *Betacoronaviruses*. Moreover, topological constraints are likely to drive the conservation of a deamidation hotspot at position 487, resembling the 481 position of SARS CoV-2, in S protein of related bat *Betacoronaviruses* that have suffered a deletion of the entire RBR

region, a critical element for hACE2 binding. Such topological constraints have been described to drive evolution of reactive residues that regulate protein function (51).

On the other hand, deamidation hotspots in the RBM cluster that are critical for hACE2 binding (positions 493 and 501) show a low conservation profile among *Sarbecovirus*. The deamidation hotspot at position 501 is in direct contact with the K353 residue in hACE2 (35) and different observations reveal its critical role in determining hACE2 affinity (33, 52). The N501Y mutation that eliminates this deamidation hotspot is a hallmark of the emerging SARS-CoV-2 variants, such as the B.1.1.7 (UK variant), the B.1.351 (South Africa variant) and the P.1 (Brazil variant) (53). In our experimental setup we did not find any significant deamidation of Asn 501 in the SARS CoV-2 RBD construct. Hotspots 493 and 501 are frequently mutated in naturally emerging variants of SARS-CoV or SARS-CoV-2, suggesting that they are susceptible to selective pressure albeit the slow deamidation half-times experimentally observed for these positions suggest that deamidation would be neutral from an evolutive standpoint.

It should be noted that experimental evidence about deamidation of Asn 501 was reported previously for this protein (27). It is expected that discrepancies may arise in using of different experimental conditions that may affect protein conformation and hence, the deamidation rate. Alternatively, deamidation at these sites might not occur in the native, well folded S protein, but might be rapidly triggered when the protein is proteolyzed during the antigenic presentation process, interfering with the host immune response (47).

Our integral kinetic model for the virion deamidation illustrates how the multimeric nature of S protein and the virion affect the accumulation of deamidate species. The model considers only two hotspots with their respective observed experimental deamidation half-times. The model shows that deamidation would not be the only molecular mechanism that reduces virus infectivity. The full deamidation time of the virion in which all of its 33 S trimers of the external surface are deamidated at both hotspots is close to 1000 days, far slower than the time required to eradicate the virus infectivity by the sole incubation at 37°C, which has been

reported to be of a couple of days (23). However partial deamidated species are readily accumulated within hours and after a day close to four trimers would host a deamidated protomer, mainly at Asn 481. This supports the possibility that deamidated hotspots can affect Spike fitness. For example, deamidation supports the co-existence of chemically diverse RBM populations (i.e, with different primary sequence) within the virion, some of which may enable the virus to evade antibody recognition while its effect in receptor binding is moderate. In this line, the E484K mutation (that coexist with the N501Y in the Brazilian SARS CoV-2 P.1 lineage (53), located nearby to Asn 481 in the RBR, has been associated with antibody escape suggesting that mutations or modifications in the RBR region may be related to an immune evasion viral mechanism (54–56). This let us speculate that conservation of deamidation hotspots in topologically equivalent positions pursues some functionality. It has been shown that deamidation of viral capsid proteins can significantly affect ligand recognition. Deamination of N373 in the Norovirus VP1 protein, which spontaneously occurs with a half-times of days yielding an isoaspartic residue, dramatically affects the recognition of histo blood group antigens, a critical step for Norovirus infection (57). On the other hand, deamidation hotspot at position 493 (Asn 479 in the SARS-CoV sequence), which is in direct contact to K31 in hACE2 (9, 35), was only observed in SARS-CoV GZ0402, which has been isolated from a handful of individuals. Unlike the most common SARS-CoV Urbani that possesses the dipeptide Asn-

Asp at positions 479-480, which is not expected to be a deamidation hotspot, the SARS-CoV GZ0402 bears a Asn-Gly hotspot. Interestingly, a deamidation hotspot is generated in antibody escape mutants of SARS-CoV recovered under the selective pressure of the R80 neutralizing monoclonal antibody. The escape variants bear a D480G mutation that transforms a ND slow-deamidating Asn into a NG deamidation hotspot (58, 59).

Overall, our data show that deamidation hotspots are enriched and conserved in the RBD of *Sarbecovirus*, and SARS-CoV-2 in particular, suggesting that deamidation at specific positions might influence viral fitness. The non-negligible accumulation of deamidated S protein within a virion during the viral life-cycle supports the possibility that deamidation mediates some functional-relevant mechanism. However, it is unclear how such a slow, spontaneous and irreversible PTM might affect viral biology. Deamidation at the conserved hotspot 481 in the RBR might induce a radical change in the charge and the backbone of the RBM enabling the virus to escape many neutralizing antibodies, or in addition, it might generate a non-canonical integrin-binding motif (50), triggering a new function in those deamidated protomers.

Beyond the identification of deamidation hotspots in the S protein of SARS-CoV-2, a previously neglected attribute with impact in the understanding of the biology of this highly infective virus, our combined rational is suitable for prioritizing the identification of any deamidation hotspots that, impacting fitness, have been evolutionarily conserved.

Experimental procedures

Deamidation estimation using NGOME-LITE.

The protein sequences of five S proteins were selected to investigate the presence of deamidation hotspots (SARS-CoV-2, SARS-CoV-URBANI, SARS-CoV-GZ0402, Bat-CoV-RaTG13 and Bat-CoV-WIV1). We used NGOME-LITE (24) to predict deamidation propensity of Asn residues by using predicted $t_{1/2}$ values. To compare sites of deamidation, the protein sequences were aligned using Clustal Omega (60) with default parameters. The residue numbering of SARS-CoV-2 S was used as reference. Conservation of deamidation prone Asn present in S protein of SARS-CoV-2 was checked by aligning S sequences from 38 Sarbecoviruses (S6 Table)

SARS CoV-2 RBD Protein expression and purification.

The SARS CoV-2 RBD (GenBank: MN908947) was expressed and purified as reported in (31). The RBD (residues 319 - 566) was expressed by transient transfection in HEK293-F containing a secretion signal, a C-terminal Sortase motif and a non-cleavable Histidine tag. Transfected cells were incubated

at 37 ° with agitation at 220 rpm and 8 % CO₂ atmosphere. Media containing secreted protein was harvested 4 days post transfection. During this "production" time, deamidation should occur adding a cumulative amount of deamidated species at zero time. All steps of the purification were performed at 4 °C, during which deamidation rate is expected to be considerably reduced. For RBD purification, an initial IMAC step was performed on a His-Trap column (all columns by GE Healthcare) using buffers A (50 mM Tris-HCl, 0.5 M NaCl, 10 mM imidazole, pH 7.4) and B (same as buffer A but with 500 mM imidazole). An additional gel filtration step was performed on a Superdex 200 Increase 10/300 GL column equilibrated with 50 mM Tris-HCl, 150 mM NaCl, pH 7.4. The protein was concentrated to 10 mg/ml and flash frozen in liquid nitrogen and stored at -80 °C until further use.

hACE2 -Protein expression and purification.

A gene encoding hACE2 residues 1-615 followed by a C-terminal HRV 3C cleavage site and a TwinStrepII-tag was cloned in the pXLG expression vector. HEK293F cells transfected with this construct were grown in a TubeSpin® bioreactor in FreeStyle293 medium for 72h at 37°C with 8% CO₂ and agitation at 180 rpm. The secreted hACE2 protein was purified from the cell culture medium using a 1 ml StrepTactin Superflow high capacity cartridge (IBA). After elution, the C-terminal TwinStrepII-tag was removed by cleavage with His6-tagged HRV 3C protease, followed by an IMAC step to remove the HRV 3C protease from the sample. Finally, the untagged hACE2 protein was injected into a Superdex200 Increase 10/300 GL size exclusion chromatography column (Cytiva) equilibrated in 50 mM Hepes, pH 7.2, 150 mM NaCl and 10% glycerol. The protein was concentrated to 1.9 mg/ml, flash-frozen in liquid nitrogen and stored at -80°C.

Chemical biotinylation of hACE2

hACE2 was chemically biotinylated using the EZ-Link NHS-PEG4-Biotin kit (Sigma). The sample was first desalted using a PD-10 gravity flow column (GE Healthcare) in 20 mM sodium phosphate buffer and 150 mM NaCl, pH 7.4. Subsequently, the sample was chemically biotinylated for 2 h on ice, using a 20-fold molar excess of biotin over the target protein. Excess biotin was removed by running the sample through a Size Exclusion Chromatography column (Superdex 200 Increase 10/300 GL). The fractions containing hACE2 were collected and concentrated to approximately 1 mg/ml. A total of 5% v/v glycerol was added before flash-freezing, and the samples were stored at -80 °C until further use.

Incubation conditions for deamidation rate determination.

Purified RBD sample were diluted to 1 mg/ml in buffer 50 mM Tris-HCl, 150 mM NaCl and 5% v/v glycerol, pH 7.4, filtered with a pore size of 0.22 µm for minimizing bacterial growth and added in sealed SafeSel vials that were further sealed with parafilm. Samples were incubated at 4 and 37°C. Independent aliquots were taken at different times for deamidation quantitation.

GluC and trypsin double digest at low pH for deamidation analysis.

Digestion was done at pH 5.7 to avoid extensive deamidation during sample processing. Protein samples were diluted 1:10 in 100mM ammonium acetate (Fluka), pH 5.7, containing 10% v/v acetonitrile (Fisher Scientific). Disulfide bonds were reduced with 1 mM DTT (Sigma) and 1 mM TCEP (Invitrogen) for 30 min at 37°C. Proteins were first digested with trypsin (Promega) for 1 h at 37°C (sample to enzyme ratio 1:20) and then with GluC (Promega) (1:40 ratio) for additional 3h at 37°C. The peptides were cleaned up using an OASIS HLB µElution Plate (Waters).

Mass spectrometry.

An UltiMate 3000 RSLC nano LC system (Dionex) fitted with a trapping cartridge (µ-Precolumn C18 PepMap 100, 5µm, 300 µm i.d. x 5 mm, 100 Å) and an analytical column (nanoEase™ M/Z HSS T3 column 75 µm x 250 mm C18, 1.8 µm, 100 Å, Waters) was used, coupled directly to an Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (Thermo). Peptide samples were loaded onto the pre-column with a constant flow rate of 30 µl/min for 6 min using 0.05% v/v trifluoroacetic acid in water. Subsequently, peptides were eluted *via* the analytical column (Solvent A: 0.1% formic acid in water) with a constant flow of 0.3 µl/min, with increasing percentage of solvent B (0.1% formic acid in acetonitrile). The peptides

were introduced into the Fusion Lumos *via* a Pico-Tip Emitter 360 μm OD x 20 μm ID; 10 μm tip (New Objective) and an applied spray voltage of 2.4 kV. The capillary temperature was set to 275°C. Full mass scan (MS1) was acquired with a mass range of 375-2000 m/z in profile mode in the orbitrap with resolution of 120000. The filling time was set at a maximum of 50 ms. Data dependent acquisition (DDA) was performed with the resolution of the Orbitrap set to 30000, with a fill time of 86 ms. A normalized collision energy of 34 was applied. MS2 data were acquired in profile mode.

Data analysis.

Raw data were searched against the Uniprot human (75,069 entries, June 2020) and the SARS-CoV-2 (14 entries, June 2020) reference proteomes using MaxQuant 1.6.14.0 (32). The default parameters were used with the following adjustments: Deamidation of Asn and Gln, were set as variable modifications, as well as oxidation of Met and acetylation of protein N-termini. No fixed modifications were enabled (only reduction, no alkylation). The minimum peptide length was adjusted to 6 or 7 and the maximum peptide mass to 8000 Da. The precursor mass tolerance of the first search was set to 20 ppm and to 4.5 ppm for main search. The mass tolerance of fragment ions was set to 20 ppm. For digestion Trypsin/P and GluC were enabled and the missed cleavages were set to 3. Label-free quantification was enabled, with Fast LFQ disabled. IBAQ calculations were enabled. The minimum number of unique peptides was set to 1. A false discovery rate (FDR) of 1% was applied on both peptide (on modified peptides separately) and protein lists. Decoy mode was set to "revert". The minimum score for unmodified peptides was 0, for modified peptides 40. The Maxquant output msms.txt was imported into Skyline 20.2.0.286 (61). Precursor ion chromatograms were extracted and the values of the peak areas were exported for further calculations. MS1 spectra were inspected manually using Thermo Xcalibur Qual browser 4.2.28.14.

Biolayer interferometry (BLI).

The binding of RBD (0 days after purification or 20 days after deamidation at 37°C) to biotinylated hACE2 was measured by biolayer interferometry (BLI) using the Octet RED96 system (FortéBio). Concentration-dependent kinetic assays were performed by loading biotinylated hACE2 on streptavidin biosensors (FortéBio), pre-equilibrated in assay buffer (PBS buffer supplemented with 0.1% (w/v) BSA and 0.02% (v/v) Tween-20) for 15 min. Prior to the association, a baseline step of 300 s was performed. Subsequently, sensors were dipped in a different well containing 200, 100, 50, 25 and 12.5 nM of RBD for 600 s followed by 900 s of dissociation time in the same buffer. All experiments were carried out at 30 °C. Data were reference-subtracted only for the association and aligned with each other with an in-house python script, using a 1:1 binding model. All figures were prepared using R and ggplot2. Two independent experiments were done.

Data availability

All data needed for the assessment or verification of the manuscript's findings is provided as supporting information. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (62) partner repository with the dataset identifier PXD028071 (peptide: IYQAGSTPCNGVE) Dataset S1 and PXD027873 (peptides: CVNFNFNGLTGTGVLTE and GFNCYFPLQSYGFQPTNGVGYQPYPYR) Dataset S2.

Supporting information

This article contains supporting information.

Acknowledgments

We would like to acknowledge the Sample Preparation and Characterization Facility at EMBL Hamburg.

Author contributions

Project design and development was done by L.G.A. R.L. performed NGOME-LITE analysis, solvent accessibility calculations and conservation analysis of the S protein. L.A., and I.E.S., developed the virion deamidation kinetic model. P.O.C, L.H.O and S.K, analyzed structural S models and performed area calculation. T.F.C and L.A.D. expressed and purified RBD SARS-CoV-2 protein. K.R. expressed and purified hACE2. Sample preparation for the MS experiments was done by L.A.D. BLI binding studies were done by L.A.D. and subsequent analysis was done by S.N. J.J.S performed the MS experiments and analyzed the data. Resources, Funding Acquisition and Supervision was carried out by L.G.A, I.E.S , C.L and M.G.A. A first draft was written by L.G.A and further edited by all authors.

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations

SARS, severe acute respiratory syndrome; CoV, coronavirus

References

1. Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G. F., and Tan, W. (2020) A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727–733
2. World Health Organization, W. H. O. (2021) Coronavirus (COVID-19) Dashboard. [online] <https://covid19.who.int> (Accessed March 15, 2021)
3. Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A. A., Haagmans, B. L., Lauber, C., Leontovich, A. M., Neuman, B. W., Penzar, D., Perlman, S., Poon, L. L. M., Samborskiy, D. V., Sidorov, I. A., Sola, I., Ziebuhr, J., and Coronaviridae Study Group of the International Committee on Taxonomy of Viruses (2020) The species Severe acute respiratory syndrome-related coronavirus : classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* **5**, 536–544
4. Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., Zheng, X.-S., Zhao, K., Chen, Q.-J., Deng, F., Liu, L.-L., Yan, B., Zhan, F.-X., Wang, Y.-Y., Xiao, G.-F., and Shi, Z.-L. (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature.* **579**, 270–273
5. Wan, Y., Shang, J., Graham, R., Baric, R. S., and Li, F. (2020) Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *J. Virol.* 10.1128/JVI.00127-20
6. Hoffmann, M., Kleine-Weber, H., and Pöhlmann, S. (2020) A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Mol. Cell.* **78**, 779-784.e5
7. Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., Graham, B. S., and McLellan, J. S. (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science.* **367**, 1260–1263
8. Wang, Y., Liu, M., and Gao, J. (2020) Enhanced receptor binding of SARS-CoV-2 through networks of hydrogen-bonding and hydrophobic interactions. *Proc. Natl. Acad. Sci.* **117**, 13967–13974
9. Li, W., Zhang, C., Sui, J., Kuhn, J. H., Moore, M. J., Luo, S., Wong, S.-K., Huang, I.-C., Xu, K., Vasilieva, N., Murakami, A., He, Y., Marasco, W. A., Guan, Y., Choe, H., and Farzan, M. (2005) Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.* **24**, 1634–1643
10. Song, H.-D., Tu, C.-C., Zhang, G.-W., Wang, S.-Y., Zheng, K., Lei, L.-C., Chen, Q.-X., Gao, Y.-W., Zhou, H.-Q., Xiang, H., Zheng, H.-J., Chern, S.-W. W., Cheng, F., Pan, C.-M., Xuan, H., Chen, S.-J., Luo, H.-M., Zhou, D.-H., Liu, Y.-F., He, J.-F., Qin, P.-Z., Li, L.-H., Ren, Y.-Q., Liang, W.-J., Yu, Y.-D., Anderson, L., Wang, M., Xu, R.-H., Wu, X.-W., Zheng, H.-Y., Chen, J.-D., Liang, G., Gao, Y., Liao, M., Fang, L., Jiang, L.-Y., Li, H., Chen, F., Di, B., He, L.-J., Lin, J.-Y., Tong, S., Kong, X., Du, L., Hao, P., Tang, H., Bernini, A., Yu, X.-J., Spiga, O., Guo, Z.-M., Pan, H.-Y., He, W.-Z., Manuguerria, J.-C., Fontanet, A., Danchin, A.,

- Niccolai, N., Li, Y.-X., Wu, C.-I., and Zhao, G.-P. (2005) Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc. Natl. Acad. Sci.* **102**, 2430–2435
11. Barnes, C. O., West, A. P., Huey-Tubman, K. E., Hoffmann, M. A. G., Sharaf, N. G., Hoffman, P. R., Koranda, N., Gristick, H. B., Gaebler, C., Muecksch, F., Lorenzi, J. C. C., Finkin, S., Hägglöf, T., Hurley, A., Millard, K. G., Weisblum, Y., Schmidt, F., Hatzioannou, T., Bieniasz, P. D., Caskey, M., Robbiani, D. F., Nussenzweig, M. C., and Bjorkman, P. J. (2020) Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies. *Cell.* **182**, 828-842.e16
 12. Liu, L., Wang, P., Nair, M. S., Yu, J., Rapp, M., Wang, Q., Luo, Y., Chan, J. F.-W., Sahi, V., Figueroa, A., Guo, X. V., Cerutti, G., Bimela, J., Gorman, J., Zhou, T., Chen, Z., Yuen, K.-Y., Kwong, P. D., Sodroski, J. G., Yin, M. T., Sheng, Z., Huang, Y., Shapiro, L., and Ho, D. D. (2020) Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature.* **584**, 450–456
 13. Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veerler, D. (2020) Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell.* **181**, 281-292.e6
 14. Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J. L., Marc, G. P., Moreira, E. D., Zerbini, C., Bailey, R., Swanson, K. A., Roychoudhury, S., Koury, K., Li, P., Kalina, W. V., Cooper, D., Robert W. Frenc, J., Hammitt, L. L., Türeci, Ö., Nell, H., Schaefer, A., Ünal, S., Tresnan, D. B., Mather, S., Dormitzer, P. R., Şahin, U., Jansen, K. U., and Gruber, W. C. (2020) Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N. Engl. J. Med.* 10.1056/NEJMoa2034577
 15. Sadoff, J., Gars, M. L., Shukarev, G., Heerwegh, D., Truyers, C., Groot, A. M. de, Stoop, J., Tete, S., Damme, W. V., Leroux-Roels, I., Berghmans, P.-J., Kimmel, M., Damme, P. V., Hoon, J. de, Smith, W., Stephenson, K. E., Rosa, S. C. D., Cohen, K. W., McElrath, M. J., Cormier, E., Scheper, G., Barouch, D. H., Hendriks, J., Struyf, F., Douoguih, M., Hoof, J. V., and Schuitemaker, H. (2021) Interim Results of a Phase 1–2a Trial of Ad26.COV2.S Covid-19 Vaccine. *N. Engl. J. Med.* 10.1056/NEJMoa2034201
 16. Reissner, K. J., and Aswad, D. W. (2003) Deamidation and isoaspartate formation in proteins: unwanted alterations or surreptitious signals? *Cell. Mol. Life Sci. CMLS.* **60**, 1281–1295
 17. Giles, A. R., Sims, J. J., Turner, K. B., Govindasamy, L., Alvira, M. R., Lock, M., and Wilson, J. M. (2018) Deamidation of Amino Acids on the Surface of Adeno-Associated Virus Capsids Leads to Charge Heterogeneity and Altered Vector Function. *Mol. Ther.* **26**, 2848–2862
 18. Geiger, T., and Clarke, S. (1987) Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. *J. Biol. Chem.* **262**, 785–794
 19. Robinson, N. E., and Robinson, A. B. (2001) Prediction of protein deamidation rates from primary and three-dimensional structure. *Proc. Natl. Acad. Sci.* **98**, 4367–4372
 20. Plotnikov, N. V., Singh, S. K., Rouse, J. C., and Kumar, S. (2017) Quantifying the Risks of Asparagine Deamidation and Aspartate Isomerization in Biopharmaceuticals by Computing Reaction Free-Energy Surfaces. *J. Phys. Chem. B.* **121**, 719–730
 21. Delmar, J. A., Wang, J., Choi, S. W., Martins, J. A., and Mikhail, J. P. (2019) Machine Learning Enables Accurate Prediction of Asparagine Deamidation Probability and Rate. *Mol. Ther. - Methods Clin. Dev.* **15**, 264–274
 22. Robinson, N. E., and Robinson, A. B. (2001) Molecular clocks. *Proc. Natl. Acad. Sci.* **98**, 944–949
 23. Chin, A. W. H., Chu, J. T. S., Perera, M. R. A., Hui, K. P. Y., Yen, H.-L., Chan, M. C. W., Peiris, M., and Poon, L. L. M. (2020) Stability of SARS-CoV-2 in different environmental conditions. *Lancet Microbe.* **1**, e10
 24. Lorenzo, J. R., Leonetti, C. O., Alonso, L. G., and Sánchez, I. E. (2020) NGOME-Lite: Proteome-wide prediction of spontaneous protein deamidation highlights differences between taxa. *Methods.* 10.1016/j.ymeth.2020.11.001
 25. Wrobel, A. G., Benton, D. J., Xu, P., Roustan, C., Martin, S. R., Rosenthal, P. B., Skehel, J. J., and Gamblin, S. J. (2020) SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nat. Struct. Mol. Biol.* **27**, 763–767
 26. Menachery, V. D., Yount, B. L., Sims, A. C., Debbink, K., Agnihothram, S. S., Gralinski, L. E., Graham, R. L., Scobey, T., Plante, J. A., Royal, S. R., Swanstrom, J., Sheahan, T. P., Pickles, R. J., Corti, D., Randell, S. H., Lanzavecchia, A., Marasco, W. A., and Baric, R. S. (2016) SARS-like WIV1-CoV poised for human emergence. *Proc. Natl. Acad. Sci.* **113**, 3048–3053
 27. Zhao, P., Praissman, J. L., Grant, O. C., Cai, Y., Xiao, T., Rosenbalm, K. E., Aoki, K., Kellman, B. P., Bridger, R., Barouch, D. H., Brindley, M. A., Lewis, N. E., Tiemeyer, M., Chen, B., Woods, R. J., and Wells, L. (2020) Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. *Cell Host Microbe.* **28**, 586-601.e6
 28. Jia, L., and Sun, Y. (2017) Protein asparagine deamidation prediction based on structures with machine learning methods. *PLOS ONE.* **12**, e0181347

29. Yan, Q., Huang, M., Lewis, M. J., and Hu, P. (2018) Structure Based Prediction of Asparagine Deamidation Propensity in Monoclonal Antibodies. *mAbs*. **10**, 901–912
30. Casalino, L., Gaieb, Z., Goldsmith, J. A., Hjorth, C. K., Dommer, A. C., Harbison, A. M., Fogarty, C. A., Barros, E. P., Taylor, B. C., McLellan, J. S., Fadda, E., and Amaro, R. E. (2020) Beyond Shielding: The Roles of Glycans in SARS-CoV-2 Spike Protein. *bioRxiv*. 10.1101/2020.06.11.146522
31. Custódio, T. F., Das, H., Sheward, D. J., Hanke, L., Pazicky, S., Pieprzyk, J., Sorgenfrei, M., Schroer, M. A., Gruzinov, A. Y., Jeffries, C. M., Graewert, M. A., Svergun, D. I., Dobrev, N., Remans, K., Seeger, M. A., McInerney, G. M., Murrell, B., Hällberg, B. M., and Löw, C. (2020) Selection, biophysical and structural analysis of synthetic nanobodies that effectively neutralize SARS-CoV-2. *Nat. Commun.* **11**, 5588
32. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372
33. Starr, T. N., Greaney, A. J., Hilton, S. K., Ellis, D., Crawford, K. H. D., Dingens, A. S., Navarro, M. J., Bowen, J. E., Tortorici, M. A., Walls, A. C., King, N. P., Veessler, D., and Bloom, J. D. (2020) Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell*. **182**, 1295-1310.e20
34. Li, F., Li, W., Farzan, M., and Harrison, S. C. (2005) Structure of SARS Coronavirus Spike Receptor-Binding Domain Complexed with Receptor. *Science*. **309**, 1864–1868
35. Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., and Wang, X. (2020) Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. **581**, 215–220
36. Lefkowitz, E. J., Dempsey, D. M., Hendrickson, R. C., Orton, R. J., Siddell, S. G., and Smith, D. B. (2018) Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res.* **46**, D708–D717
37. Wells, H. L., Letko, M., Lasso, G., Ssebidde, B., Nziza, J., Byarugaba, D. K., Navarrete-Macias, I., Liang, E., Cranfield, M., Han, B. A., Tingley, M. W., Diuk-Wasser, M., Goldstein, T., Johnson, C. K., Mazet, J., Chandran, K., Munster, V. J., Gilardi, K., and Anthony, S. J. (2021) The evolutionary history of ACE2 usage within the coronavirus subgenus Sarbecovirus. *bioRxiv*. 10.1101/2020.07.07.190546
38. Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., and Li, F. (2020) Structural basis of receptor recognition by SARS-CoV-2. *Nature*. **581**, 221–224
39. Zhang, S., Qiao, S., Yu, J., Zeng, J., Shan, S., Tian, L., Lan, J., Zhang, L., and Wang, X. (2021) Bat and pangolin coronavirus spike glycoprotein structures provide insights into SARS-CoV-2 evolution. *Nat. Commun.* **12**, 1607
40. Conceicao, C., Thakur, N., Human, S., Kelly, J. T., Logan, L., Bialy, D., Bhat, S., Stevenson-Leggett, P., Zagrajek, A. K., Hollinghurst, P., Varga, M., Tsirigoti, C., Tully, M., Chiu, C., Moffat, K., Silesian, A. P., Hammond, J. A., Maier, H. J., Bickerton, E., Shelton, H., Dietrich, I., Graham, S. C., and Bailey, D. (2020) The SARS-CoV-2 Spike protein has a broad tropism for mammalian ACE2 proteins. *PLOS Biol.* **18**, e3001016
41. World Health Organization, W. H. O. (2021) SARS-CoV-2 Variants. *WHO - Dis. Outbreak News*. [online] <https://www.who.int/csr/don/31-december-2020-sars-cov2-variants/en/> (Accessed December 31, 2020)
42. Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., and Kummer, U. (2006) COPASI—a COMplex PATHway Simulator. *Bioinformatics*. **22**, 3067–3074
43. Ke, Z., Oton, J., Qu, K., Cortese, M., Zila, V., McKeane, L., Nakane, T., Zivanov, J., Neufeldt, C. J., Cerikan, B., Lu, J. M., Peukes, J., Xiong, X., Kräusslich, H.-G., Scheres, S. H. W., Bartenschlager, R., and Briggs, J. A. G. (2020) Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature*. **588**, 498–502
44. Klein, S., Cortese, M., Winter, S. L., Wachsmuth-Melm, M., Neufeldt, C. J., Cerikan, B., Stanifer, M. L., Boulant, S., Bartenschlager, R., and Chlanda, P. (2020) SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *bioRxiv*. 10.1101/2020.06.23.167064
45. Goodsell, D. S., Voigt, M., Zardecki, C., and Burley, S. K. (2020) Integrative illustration for coronavirus outreach. *PLOS Biol.* **18**, e3000815
46. Cirrito, T. P., Pu, Z., Deck, M. B., and Unanue, E. R. (2001) Deamidation of Asparagine in a Major Histocompatibility Complex–Bound Peptide Affects T Cell Recognition but Does Not Explain Type B Reactivity. *J. Exp. Med.* **194**, 1165–1170
47. Moss, C. X., Matthews, S. P., Lamont, D. J., and Watts, C. (2005) Asparagine Deamidation Perturbs Antigen Presentation on Class II Major Histocompatibility Complex Molecules *. *J. Biol. Chem.* **280**, 18498–18503
48. Doyle, H. A., Gee, R. J., and Mamula, M. J. (2007) Altered immunogenicity of isoaspartate containing proteins. *Autoimmunity*. **40**, 131–137
49. Deverman, B. E., Cook, B. L., Manson, S. R., Niederhoff, R. A., Langer, E. M., Rosová, I., Kulans, L. A.,

- Fu, X., Weinberg, J. S., Heinecke, J. W., Roth, K. A., and Weintraub, S. J. (2002) Bcl-xL Deamidation Is a Critical Switch in the Regulation of the Response to DNA Damage. *Cell*. **111**, 51–62
50. Curnis, F., Longhi, R., Crippa, L., Cattaneo, A., Dondossola, E., Bachi, A., and Corti, A. (2006) Spontaneous Formation of L-Isoaspartate and Gain of Function in Fibronectin *. *J. Biol. Chem.* **281**, 36466–36476
 51. Alvarez-Paggi, D., Lorenzo, J. R., Camporeale, G., Montero, L., Sánchez, I. E., de Prat Gay, G., and Alonso, L. G. (2019) Topology Dictates Evolution of Regulatory Cysteines in a Family of Viral Oncoproteins. *Mol. Biol. Evol.* **36**, 1521–1532
 52. Yi, C., Sun, X., Ye, J., Ding, L., Liu, M., Yang, Z., Lu, X., Zhang, Y., Ma, L., Gu, W., Qu, A., Xu, J., Shi, Z., Ling, Z., and Sun, B. (2020) Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies. *Cell. Mol. Immunol.* **17**, 621–630
 53. Galloway, S. E. (2021) Emergence of SARS-CoV-2 B.1.1.7 Lineage — United States, December 29, 2020–January 12, 2021. *MMWR Morb. Mortal. Wkly. Rep.* 10.15585/mmwr.mm7003e2
 54. Baum, A., Fulton, B. O., Wloga, E., Copin, R., Pascal, K. E., Russo, V., Giordano, S., Lanza, K., Negron, N., Ni, M., Wei, Y., Atwal, G. S., Murphy, A. J., Stahl, N., Yancopoulos, G. D., and Kyratsous, C. A. (2020) Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science*. **369**, 1014–1018
 55. Ku, Z., Xie, X., Davidson, E., Ye, X., Su, H., Menachery, V. D., Li, Y., Yuan, Z., Zhang, X., Muruato, A. E., Escuer, A. G., Tyrell, B., Doolan, K., Doranz, B. J., Wrapp, D., Bates, P. F., McLellan, J. S., Weiss, S. R., Zhang, N., Shi, P.-Y., and An, Z. (2021) Molecular determinants and mechanism for antibody cocktail preventing SARS-CoV-2 escape. *Nat. Commun.* **12**, 469
 56. Xie, X., Liu, Y., Liu, J., Zhang, X., Zou, J., Fontes-Garfias, C. R., Xia, H., Swanson, K. A., Cutler, M., Cooper, D., Menachery, V. D., Weaver, S. C., Dormitzer, P. R., and Shi, P.-Y. (2021) Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. *Nat. Med.* 10.1038/s41591-021-01270-4
 57. Mallagaray, A., Creutzmacher, R., Dülfer, J., Mayer, P. H. O., Grimm, L. L., Orduña, J. M., Trabjerg, E., Stehle, T., Rand, K. D., Blaum, B. S., Uetrecht, C., and Peters, T. (2019) A post-translational modification of human Norovirus capsid protein attenuates glycan binding. *Nat. Commun.* **10**, 1320
 58. Sui, J., Aird, D. R., Tamin, A., Murakami, A., Yan, M., Yammanuru, A., Jing, H., Kan, B., Liu, X., Zhu, Q., Yuan, Q., Adams, G. P., Bellini, W. J., Xu, J., Anderson, L. J., and Marasco, W. A. (2008) Broadening of Neutralization Activity to Directly Block a Dominant Antibody-Driven SARS-Coronavirus Evolution Pathway. *PLOS Pathog.* **4**, e1000197
 59. Sui, J., Deming, M., Rockx, B., Liddington, R. C., Zhu, Q. K., Baric, R. S., and Marasco, W. A. (2014) Effects of Human Anti-Spike Protein Receptor Binding Domain Antibodies on Severe Acute Respiratory Syndrome Coronavirus Neutralization Escape and Fitness. *J. Virol.* **88**, 13769–13780
 60. Sievers, F., and Higgins, D. G. (2018) Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* **27**, 135–145
 61. Schilling, B., Rardin, M. J., MacLean, B. X., Zawadzka, A. M., Frewen, B. E., Cusack, M. P., Sorensen, D. J., Bereman, M. S., Jing, E., Wu, C. C., Verdin, E., Kahn, C. R., MacCoss, M. J., and Gibson, B. W. (2012) Platform-independent and Label-free Quantitation of Proteomic Data Using MS1 Extracted Ion Chromatograms in Skyline: APPLICATION TO PROTEIN ACETYLATION AND PHOSPHORYLATION *. *Mol. Cell. Proteomics.* **11**, 202–214
 62. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, Ş., Tiwary, S., Cox, J., Audain, E., Walzer, M., Jarnuczak, A. F., Ternent, T., Brazma, A., and Vizcaíno, J. A. (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450

Footnotes

Funding and additional information

We acknowledge funding from Agencia Nacional de Promoción Científica y Tecnológica (PICT 2015-1213 to I.E.S.) and Consejo Nacional de Investigaciones Científicas y Técnicas (I.E.S., L.G.A., L.H.O and S.K are CONICET career investigators and R.L. is a member of the CONICET Support Staff for Research and Development Career). L.A.D was supported by the EMBL Interdisciplinary Postdoc

Program (EIPOD) under Marie Curie COFUND actions (Grant Agreement 664726). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of manuscript preparation.

Journal Pre-proof

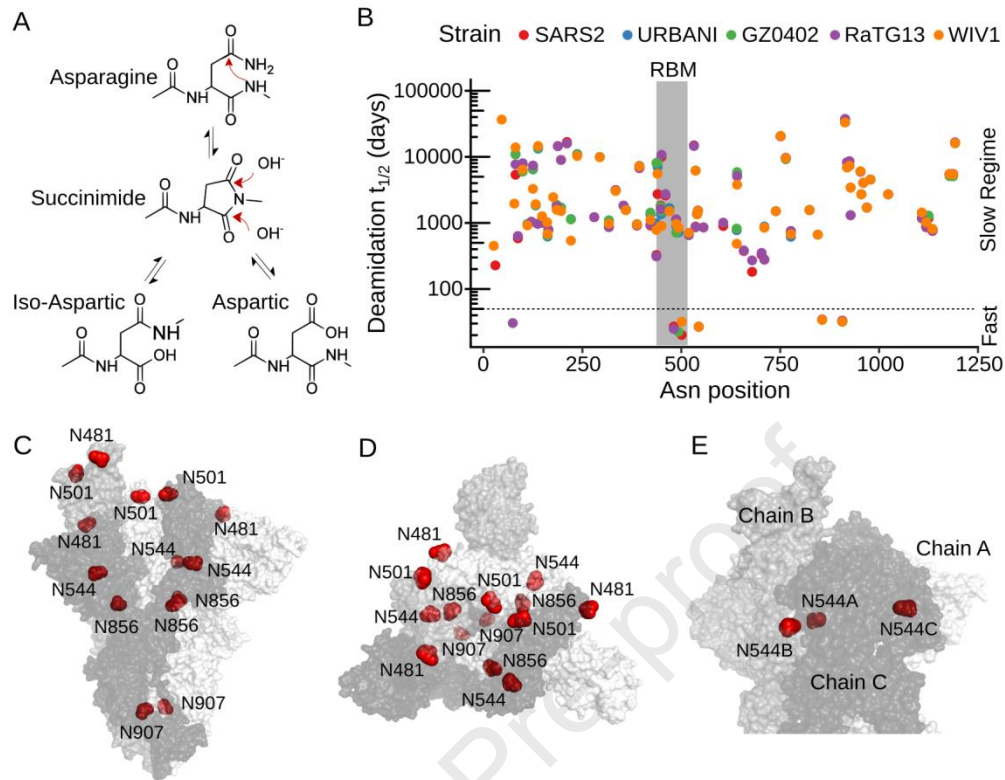


Fig 1. Deamidation profile of S from *Betacoronaviruses*.

(A) Schematic representation of the deamidation reaction mechanism. (B) Deamidation profile of S proteins from a group of *Betacoronaviruses* obtained with NGOME-LITE; only non-glycosylable asparagines are included. Deamidation half-times are shown on logarithmic scale. The Asn positions are referred to the SARS-CoV-2 sequence. (C) Deamidation hotspots on the SARS-CoV-2 S protein are shown in red. Protomers are colored in different gray tones. The PDB 6zgg (25) was used for this analysis, which contains a furin-cleaved SARS-CoV-2 S trimer, with two RBDs in the down conformation (chains A and C) and one in the up conformation (chain B). (D) Top view of (C), showing exposed deamidation hotspots at the RBD (Asn 481 and 501). (E) Zoom of the exposed Asn 544 residues.

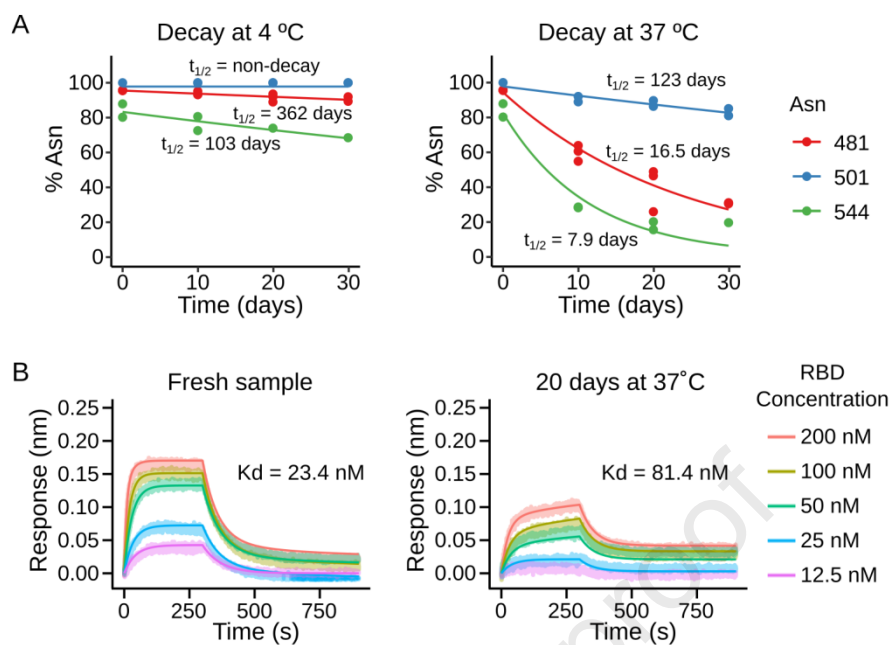


Fig 2. Experimental deamidation kinetic and affinity to hACE2 of an aged RBD sample.

(A) Left: Time-decay of asparagine-containing peptides for the 481, 501 and 544 hotspots at 4 °C. The lines are fit to an exponential function with initial amplitude of 100% and endpoint of 0%. Right: Time-decay of asparagine-containing peptides for the 481, 501 and 544 hotspots at 37 °C. The lines are fitted to an exponential decay. (B) Left: BLI response curves of a fresh RBD sample to an immobilized hACE2. Right: BLI response curve of an aged RBD sample (20 days at 37 °C).

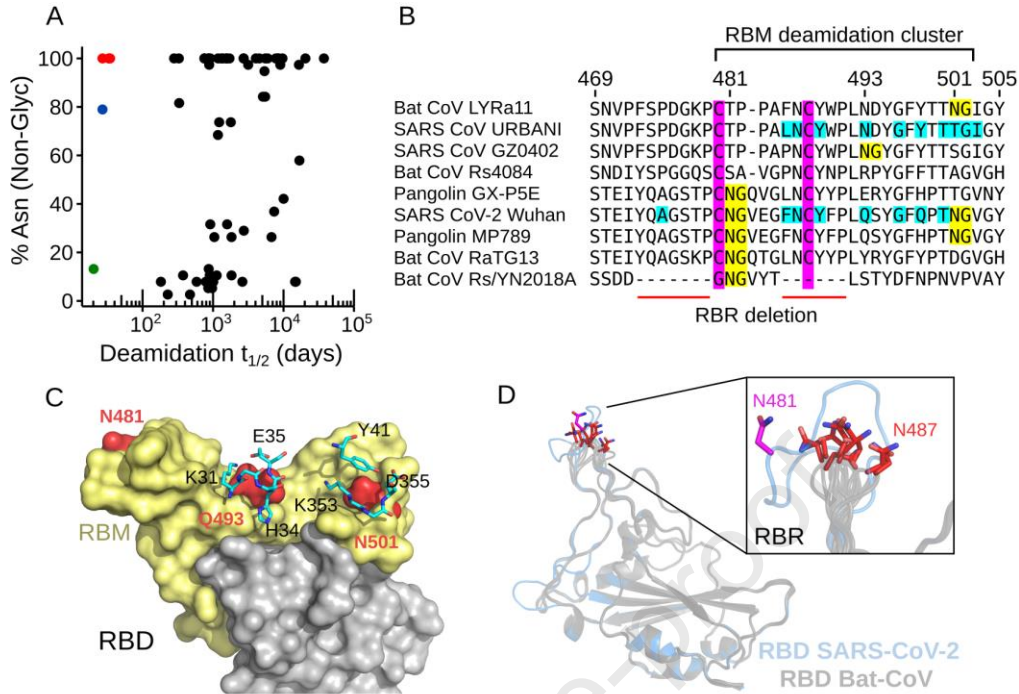


Fig 3. Conservation pattern of deamidation hotspots at the RBM of Sarbecoviruses.

(A) Conservation of asparagines in SARS-CoV-2 showed as the percentage of Asn present in 38 *Sarbecoviruses* vs deamidation half-times (Dataset S3). Hotspots 544, 856 and 907, red dots; hotspot 481, blue dot and hotspot 501, green dot. (B) Alignment of different S proteins from selected *Sarbecoviruses*. The RBM region is indicated and the deamidation hotspots are highlighted in yellow. Residues of SARS-CoV and SARS-CoV-2 located closer than 5Å from hACE2 residues, are highlighted in cyan. (C) Location of the deamidation hotspots 481, 493 and 501 at the surface of SARS-CoV-2 RBD (pdb: 6M0J). Deamidation hotspots are shown in red, RBM (residues 438-506) is shown in ocher and the core of the RBD is depicted in gray. hACE2 residues that directly interact with the RBM deamidation hotspots 493 (K31, H34 and E35) and 501 (Y41, K353 and D355) are shown as sticks. In SARS-CoV-2, the position 493 is a glutamine. (D) Superposition of RBD from SARS-CoV-2 (pdb 6M0J, pale blue) and an ensemble of modeled structures of RBD from bat-CoV Rs/YN2018A (gray). Asn 481 residue in SARS-CoV-2 RBD is shown in magenta sticks and the Asn residues corresponding to the predicted 487 deamidation hotspot in Rs/YN2018A are drawn in red sticks. The inset details the receptor binding ridge (RBR) region.

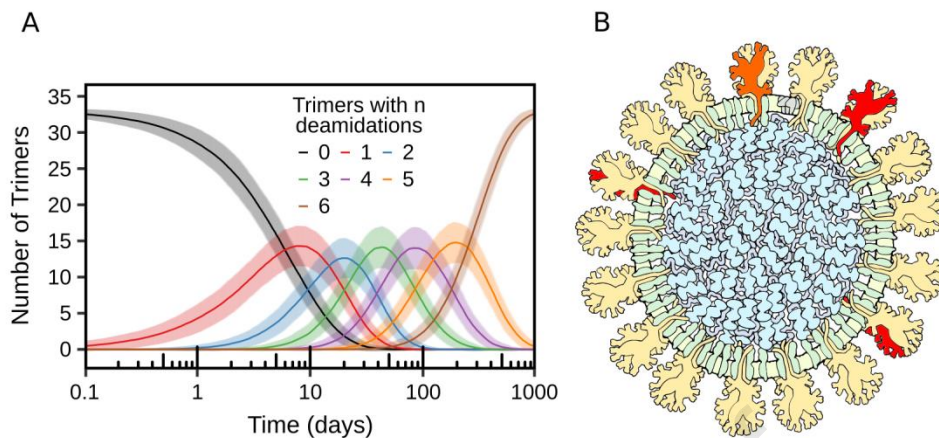


Fig 4. Spike protein hotspot deamidation in the context of the SARS-CoV2 virion at 37 °C.

(A) Simulated time course of the number of trimers with 0 to 6 deamidation events at sites Asn 481 and Asn 501, using the deamidation half-times from Fig 2 (Fig S3, for simulation details). We report the average and standard deviation of 1000 simulations using the Gillespie algorithm (Fig S4 for full results and Table S8 for the results at several time points of interest). (B) Visualization of S protein deamidation in the SARS-CoV2 virion 48 h after synthesis. A cross section of the virus is shown (see (23) for details), with membrane and membrane-bound viral proteins M and E in green and the viral genome and associated nucleocapsid proteins N in blue. A total of 17 of the 33 spike trimers are shown, colored according to the simulation results (Table S8). Yellow: undeamidated spike monomers. Red: spike monomers deamidated at Asn 481. Orange: spike monomers deamidated at Asn 501.

