Enhanced ER proteostasis and temperature differentially impact the mutational tolerance of influenza hemagglutinin

Angela M. Phillips;¹ Michael B. Doud;⁵,⁶ Luna O. Gonzalez;² Vincent Butty;³ Yu-Shan Lin;⁴ Jesse D. Bloom;⁵,⁶ Matthew D. Shoulders¹

¹Department of Chemistry, ²Department of Mathematics, ³BioMicro Center, Massachusetts Institute of Technology, Cambridge, United States ⁴Department of Chemistry, Tufts University, Medford, United States ⁵Fred Hutchinson Cancer Research Center, Seattle, United States; ⁶Department of Genome Sciences, University of Washington, Seattle, United States

*To whom correspondence may be addressed:

Matthew D. Shoulders
Department of Chemistry
Massachusetts Institute of Technology
77 Massachusetts Avenue
Building 16, Room 573A
Cambridge, MA 02139
Phone: (617) 452-3525
Fax: (617) 324-0505
Email: mshoulde@mit.edu
Herein, we systematically and quantitatively evaluate whether endoplasmic reticulum (ER) proteostasis factors impact the mutational tolerance of secretory pathway proteins. We focus on influenza hemagglutinin (HA), a viral membrane protein that folds in the host's ER via a complex pathway. By integrating chemical methods to modulate ER proteostasis with deep mutational scanning to assess mutational tolerance, we discover that upregulation of ER proteostasis factors broadly enhances HA mutational tolerance across diverse structural elements. Remarkably, this proteostasis network-enhanced mutational tolerance occurs at the same sites where mutational tolerance is most reduced by propagation at fever-like temperature. These findings have important implications for influenza evolution, because influenza immune escape is contingent on HA possessing sufficient mutational tolerance to evade antibodies while maintaining the capacity to fold and function. More broadly, this work provides the first experimental evidence that ER proteostasis mechanisms define the mutational tolerance and, therefore, the evolution of secretory pathway proteins.
INTRODUCTION

Protein evolution is necessarily constrained by the inherent biophysical properties of polypeptide sequences (1-6). Gene variants that encode proteins unable to properly fold are selected against (Figure 1A) (7), irrespective of any theoretically beneficial new function the encoded protein may otherwise have acquired (1-6). Importantly, the consequences of amino acid substitutions for protein folding depend not just on inherent biophysical properties, but also on the environment in which the nascent chain attempts to fold. In cells, this environment is defined most importantly by the proteostasis network, which consists of a subcellular compartment-specific array of chaperones and quality control factors (8-11). These proteostasis factors assist client proteins in navigating complex folding landscapes (2, 12), and are thus theoretically well-positioned to modulate the mutational landscape accessible to evolving proteins (Figure 1B).

The proteostasis network consists of ~1000 factors involved in client protein folding, modification, trafficking, and quality control. Yet, to date, virtually all studies experimentally examining evolutionary consequences of eukaryotic proteostasis network factors have focused on a single chaperone – HSP90 (13-19). These focused studies have shown that HSP90 can (1) buffer the phenotypic effects of standing genetic diversity (14, 15, 18, 19) and (2) potentiate the phenotypic effects of new mutations (16, 17, 19), perhaps by facilitating protein folding and minimizing aggregation (17). More recently, we leveraged chemical biology tools to demonstrate that cytosolic proteostasis factors beyond HSP90 can also modulate protein evolutionary trajectories and mutational tolerance (20).

The potentially critical impact of the extensive endoplasmic reticulum (ER) proteostasis machinery on the mutational tolerance and evolution of secretory pathway client proteins has remained unexplored. This gap in knowledge is important because roughly 1/3 of the proteome folds in the ER (9). Moreover, thousands of mutations in ER client proteins cause hundreds of currently incurable diseases, ranging from neurodegenerative pathologies to the lysosomal storage disorders and the collagenopathies (8). Additionally, viral membrane proteins often fold
in the ER and must rapidly evolve to escape host immune factors (21-23). Thus, understanding how ER chaperones influence the ability of client proteins to tolerate mutations and adapt to changing environments is essential as we work to design effective therapies for these diseases.

Influenza hemagglutinin (HA) is a uniquely attractive model protein for systematically and quantitatively evaluating whether and how ER proteostasis mechanisms impact client protein mutational tolerance. The folding of HA is perhaps as well-delineated as for any other membrane protein, and HA interacts extensively with components of the host cell’s ER proteostasis network (21, 22, 24-33). Co-translationally, HA is heavily N-glycosylated and engages the ER’s lectin chaperones, calnexin and calreticulin, which increase HA’s folding efficiency and prevent misfolding (21, 22, 28). During the folding process, HA also interacts with protein disulfide isomerases to form its six disulfide bonds, as well as with binding immunoglobulin protein (BiP) and glucose-regulated protein 94 (GRP94), which are the ER homologs of HSP70 and HSP90, respectively (22, 27). Together, these host ER chaperones and folding enzymes assist HA monomer folding and prevent premature trimerization, ensuring that HA reaches its functional conformation or otherwise is targeted for ER-associated degradation (8, 11, 26).

Amino acid substitutions that impair HA stability, folding, or trafficking reduce HA fitness (3, 28, 31). ER proteostasis factors that assist HA folding and trafficking are thus poised to modulate HA mutational tolerance (Figure 1B). Because HA variant fitness is directly coupled to viral replication via HA’s role in host cell binding and fusion (24, 30), we believed that we could apply a deep mutational scanning strategy (34) to quantitatively evaluate how ER proteostasis impacts the fitness of all viable single amino acid HA variants. Specifically, we aimed to quantify HA variant fitness by sequencing HA mutant viral libraries after selection in modified ER proteostasis environments, an approach that is substantially more sensitive and higher-throughput than the alternative functional assays employed for deep mutational scanning of non-viral secretory pathway clients (34).
In summary, HA is an attractive model secretory protein because HA folding is well-characterized, HA interacts extensively with host ER proteostasis factors, and HA mutational tolerance can be comprehensively quantified by pairing deep mutational scanning with sequencing. Moreover, understanding factors that both constrain and enhance HA mutational tolerance has direct therapeutic relevance. HA is the primary target of neutralizing antibodies (23), and influenza immune escape is thus contingent on HA possessing sufficient mutational tolerance to acquire antibody resistance while still maintaining the capacity to fold and function. The impact of host chaperones on viral evolutionary trajectories has only recently been realized (20), and it remains unknown whether, and to what degree, host proteostasis mechanisms impact viral protein mutational tolerance.

Below, we integrate deep mutational scanning with small molecule control of ER proteostasis environments to test the hypothesis that upregulating host ER proteostasis factors mitigates the fitness consequences of otherwise biophysically deleterious amino acid substitutions in HA (Figure 1B). We examine HA mutational tolerance in cells with basal versus upregulated levels of ER proteostasis factors, both at normal body temperature and at an elevated, fever-like temperature. We find that upregulation of the ER proteostasis machinery broadly enhances HA mutational tolerance, while increased temperature generally reduces HA mutational tolerance. Remarkably, the same HA sites where variants are most commonly temperature-sensitive, and therefore most likely to be biophysically problematic, are also the sites where variants benefit most from upregulation of the host’s ER proteostasis machinery. Thus, enhanced host ER proteostasis mechanisms and increased temperature modulate HA mutational tolerance in opposite directions, a finding with important implications for influenza evolution. More broadly, this work provides the first experimental evidence that the composition of the ER proteostasis network can profoundly impact the mutational tolerance and evolution of secretory pathway clients.
RESULTS

Modulating ER proteostasis during influenza infection

The composition of the ER proteostasis network is regulated by a stress-responsive signaling pathway termed the unfolded protein response (UPR). Under ER stress conditions, the UPR is activated and the IRE1 and ATF6 signaling cascades upregulate distinct but overlapping sets of ER proteostasis network components via the XBP1s and ATF6f transcription factors, respectively (11). Here, we profiled HA mutational tolerance at both a normal (37 °C) and a fever-like (39 °C) temperature in distinctive host ER proteostasis environments. ER proteostasis was modulated by inducing either XBP1s alone or in concert with ATF6f in a small molecule-mediated, stress-independent manner in HEK293ATF6f/XBP1s cells using previously characterized chemical genetic methods (11). In these cells, activation of tetracycline repressor-regulated XBP1s by the small molecule doxycycline induces the transcriptional response downstream of IRE1, remodeling the ER proteostasis environment by enhancing transcript levels of ~180 ER chaperones and quality control factors (9, 11). Similarly, induction of destabilized domain-regulated ATF6f by the small molecule trimethoprim in HEK293ATF6f/XBP1s cells increases transcript expression of ~40 ER chaperones and quality control factors that are distinct from but partially overlapping with those regulated by XBP1s (9, 11). Concomitant induction of ATF6f and XBP1s creates a third, distinct ER proteostasis environment, increasing expression of ~350 genes (11). We specifically assessed HA mutational tolerance in a basal ER proteostasis environment and in the two most significantly remodeled proteostasis environments, activating XBP1s alone or simultaneously activating XBP1s and ATF6f.

The media composition and treatment regimen for influenza propagation differed from the conditions used in the original characterization of XBP1s and ATF6f/XBP1s activation in these cells, which also was performed only at 37 °C (11). Hence, before quantifying HA mutational tolerance in these distinctive cellular settings, we first comprehensively characterized each environment in the absence of an influenza infection using RNA-seq. We examined...
changes in ER and cytosolic proteostasis factor transcript expression (Figure 2), as well as global transcriptional changes to assess any off-target transcriptional effects (Figure 2-figure supplement 1). Essentially all transcripts upregulated in the original characterization of these cells (11) were also upregulated in response to our modified treatment regimen (Figure 2–figure supplement 2). Likely owing to the increased sensitivity of RNA-Seq compared to whole-genome arrays (35), we also observed upregulation of 201 transcripts upon XBP1s induction and 268 transcripts upon ATF6f/XBP1s induction not observed in the previous analysis. Still, as expected (11), transcripts upregulated upon XBP1s and ATF6f/XBP1s induction largely corresponded to secretory pathway and ER stress response genes (Figure 2–figure supplement 1 and Figure 2–source data 2).

Specifically, we found that XBP1s and ATF6f/XBP1s activation in HEK293ATF6f/XBP1s cells selectively enhanced transcript expression of numerous ER proteostasis factors, including many known to interact with HA, such as calnexin (CANX) (21, 22, 28), calreticulin (CALR) (21, 22, 28), and members of the protein disulfide isomerase family (PDI) (Figure 2A, top) (22). Concomitant induction of XBP1s and ATF6f revealed increased expression of ATF6f and ATF6f/XBP1s targets, including BiP (HSPA5) and GRP94 (HSP90B1), respectively (Figure 2A, top). Stress-independent activation of XBP1s or ATF6f/XBP1s did not significantly impact transcript levels of cytosolic protein folding factors at either 37 °C or 39 °C, highlighting the selectivity of this approach for modulating ER proteostasis (Figures 2A–B, bottom; Figure 2–figure supplements 1A–D). Increasing the host cell temperature to 39 °C induced significantly fewer transcriptional changes than did activating XBP1s or ATF6f/XBP1s. Increased temperature did result in mild induction of the heat shock response, as expected (illustrated by elevated HSP70 (HSPA1A) levels in Figure 2C and Figure 2–figure supplement 1E).

We next assessed whether induction of the host cell’s ER proteostasis network and increased temperature altered host cell viability or prevented influenza propagation. First, we activated XBP1s alone or with ATF6f at both 37 °C and 39 °C in the absence of influenza, and eval-
uated cell viability 72 hours later, corresponding to the duration of our intended drug pre-
treatment and influenza infection. Consistent with previous studies (11), these perturbations
were not cytotoxic *(Figure 2–figure supplement 3A)*. Next, we evaluated whether influenza
could propagate in these remodeled host cells. Briefly, we infected cells in each pre-activated
selection environment with wild-type influenza A/WSN/1933 (the strain used throughout this
work; hereafter referred to as influenza) and titrated the viral supernatant 48 hours post-infection
using a TCID<sub>50</sub> assay. Influenza growth was moderately attenuated at increased temperature
*(Figure 2–figure supplement 3B)*, but was still sufficient for performing selections and evaluat-
ing HA mutational tolerance by sequencing.

Finally, we evaluated whether our methods for perturbing ER proteostasis were func-
tional during an influenza infection cycle. Using qPCR, we found that XBP1s and ATF6f target
genes were still inducible by our chemical genetic methods in the presence of influenza, with
only slightly attenuated upregulation relative to a mock-infection *(Figure 2–figure supplement
3C)*. This modest reduction in chaperone transcript expression was likely caused by global sup-
pression of host transcription by influenza (36). However, overall our tools for modulating ER
proteostasis factors maintained efficacy during the course of an influenza infection.

**Deep mutational scanning of HA in modulated ER proteostasis environments**

To systematically profile HA mutational tolerance in modulated ER proteostasis environments,
we employed deep mutational scanning (DMS), a method that couples saturating mutagenesis
with fitness measurements, thereby efficiently sampling amino acid sequence space (23, 34,
37). We used biological triplicate HA viral mutant libraries that were previously generated from
independent HA plasmid libraries using cells with basal levels of proteostasis factors at 37 °C
(37). The resulting HA mutant viral libraries contained nearly all viable single amino acid HA var-
iants (37). Variant sensitivity to ER proteostasis mechanisms was tested by simultaneously
competing variants within each viral library in host cells with modulated ER proteostasis environments. Each of these competitions was carried out at both 37 °C and 39 °C, to impose additional biophysical selection pressure on HA and to mimic elevated host body temperature during an influenza infection (Figure 3A).

We designed these DMS experiments to fully sample the library diversity whilst minimizing co-infection and genetic hitchhiking. To achieve this goal, we infected cells in each pre-activated selection environment with $10^6$ virions at an MOI of 0.01 infectious virion/cell. Following a 48-hour infection, or about four replication cycles, we harvested viral RNA and prepared subamplicon sequencing libraries for accurate variant frequency quantitation (37). We then quantified the differential selection (diffsel) for each mutation as the logarithm of its enrichment in the selection condition relative to that of the mock-selection condition (e.g., Basal 39 °C relative to Basal 37 °C) (23, 38). Hence, the wild-type residue at each site has a diffsel of zero, variants with positive diffsel values are more fit in the selection condition than the mock condition relative to wild-type, and variants with negative diffsel values are less fit in the selection condition than the mock condition relative to wild-type. The diffsel for variants across HA can be visualized with sequence logo plots, where the size of the amino acid letter abbreviations are proportional to the magnitude of the differential selection for that variant (Figure 3B) (23).

To decipher selection from the inherent experimental noise of batch competitions, we filtered the DMS data to include only variants with robust mutational differential selection across biological triplicates. Specifically, we limited our analyses to variants present in each pre-selection replicate viral library (Figure 3C—Triplicate mutant libraries) that exhibited differential selection of the same direction across biological triplicates. These criteria reduced experimental noise owing to differences in replicate library composition, as well as uneven sampling and biased PCR amplification of low frequency variants. This filtering predominantly excluded variants that were negligibly affected by the selection, which often had slightly positive diffsel values in one replicate versus slightly negative diffsel values in another replicate. These criteria also ex-
cluded strongly enriched or depleted variants that were not present in triplicate pre-selection libraries and, less frequently, strongly enriched or depleted variants that did not behave reproducibly across triplicate selections. Notably, the unfiltered biological replicate diffsel values were strongly correlated across HA sites, with correlation coefficients ranging from $R = 0.67–0.78$ (Figure 3–source data 1 and Dataset 1).

The filtered DMS data revealed that many HA variants were substantially depleted upon increasing the temperature (Figure 4A; Figure 4–figure supplement 1). In contrast, many variants were moderately enriched upon induction of ER proteostasis machinery (Figure 4B–E; Figure 4–figure supplements 2–5). This result was consistent with our hypothesis, as we anticipated that increasing the temperature would create a biophysically restrictive environment, reducing the fitness of HA variants, whereas increasing levels of ER proteostasis factors would create a biophysically permissive environment, enhancing the fitness of HA variants (Figure 1B).

To globally examine the relative enrichment of HA variants, we plotted HA mutant diffsel values and examined the deviation of the distribution mean from zero, which indicates no differential selection (Figure 5A). This analysis confirmed that increased temperature indeed generally reduced mutant HA viral growth, resulting in depletion (i.e. diffsel < 0) of 597 variants versus enrichment (i.e. diffsel > 0) of only 361 variants, relative to wild-type HA (Figure 5A). Variant depletion at elevated temperature could be explained by reduced variant folding rates and thus slower variant viral growth, or perhaps by increased degradation. Although we do not observe upregulation of ER-associated degradation factors at elevated temperature (Figure 2C and Figure 2-figure supplement 1E), it is certainly possible that HA variants engage degradation factors more at an elevated temperature.

In contrast, induction of XBP1s alone or with ATF6f generally enhanced mutant HA viral growth at both 37 °C and 39 °C, resulting in an overall enrichment of HA variants relative to wild-type HA (Figure 5A). The enhanced variant growth upon XBP1s induction alone was quite simi-
lar to that of simultaneous induction of XBP1s and ATF6f (Figure 5–figure supplement 1 and Figure 5–source data 1), and is thus either predominantly caused by XBP1s, or by XBP1s-regulated factors that are redundant with those regulated by ATF6f. In either case, variant enrichment upon induction of XBP1s or ATF6f/XBP1s is likely explained by upregulation of ER protein folding factors (Figures 2A–B), while the lesser amount of variant depletion observed upon induction of XBP1s or ATF6f/XBP1s may be attributed to upregulation of ER quality control and degradation factors (Figures 2A–B). Importantly, we note that the enrichment and depletion of HA variants may also be caused by indirect effects of inducing XBP1s or ATF6f/XBP1s and increasing temperature. For example, these perturbations could potentially impact the protein levels and activities of other host proteins that engage HA, or perhaps differentially modulate viral growth kinetics.

Regardless of their precise origins, the generally opposing effects of increased temperature and upregulated proteostasis machinery on mutant HA viral growth motivated examination of these effects on individual HA variants. To this end, we determined the correlation between the average diffsel values for increased temperature and upregulated ER proteostasis machinery for individual HA variants (Figures 5B–C). This analysis revealed significant negative correlation between these selection conditions, demonstrating that the HA variants most depleted upon increased temperature corresponded to those most enriched upon induction of ER proteostasis factors.

**Examination of variants and sites impacted most by host proteostasis and temperature**

The opposing fitness effects of proteostasis mechanisms and temperature in the DMS data suggested that temperature-sensitive HA variants were most likely to be rescued by ER proteostasis factors. The potential implications of these opposing selection forces on HA motivated us to more rigorously evaluate this behavior for individual HA variants. DMS batch competitions are inherently noisy, largely due to uneven sampling of library variants (37). Hence, we performed
pairwise competitions between wild-type influenza and influenza encoding differentially selected variants identified in the DMS. These competitions were performed in cells with either basal or elevated levels of ER proteostasis factors, accessed by inducing XBP1s alone or in combination with ATF6f, and at 37 °C and 39 °C (Figure 6A). We completed these pairwise competitions in biological triplicate for nine variants that in the DMS exhibited the most divergent behavior upon induction of XBP1s or ATF6f/XBP1s compared to increased temperature (A90D, K136N, S153E, K169G, P175G, K176F, P228A, L456D, and L456R; based on sequential numbering of WSN HA sites). We also included three synonymous negative controls (P175Psyn, P228Psyn, L456Lsyn), for which we mutated the wild-type codon to a synonymous codon that was not differentially selected in the DMS batch competition.

Comparison of the differential selection upon increased temperature with that upon XBP1s or ATF6f/XBP1s induction in the pairwise competitions recapitulated the divergent behavior observed in DMS for several individual HA variants. In particular, of the nine non-synonymous variants tested, seven variants exhibited significantly different behavior upon increased temperature versus XBP1s or ATF6f/XBP1s induction (Figure 6B—black; Figure 6–figure supplement 1A). This high rate of reproducibility lends substantial credence to the overall DMS data set, and we note that some discrepancy between the DMS and pairwise competitions (Figure 6B—red; Figure 6–figure supplement 1B) is expected given the difference in wild-type frequencies between the DMS and pairwise competitions and the inherent false discovery rate of high-throughput screens (37). As in the DMS (Figure 4), the pairwise diffsel upon increased temperature was overall more substantial than that upon XBP1s or ATF6f/XBP1s induction, though XBP1s or ATF6f/XBP1s induction did significantly impact the growth of K136N, P228A, and L456D (Figure 6–figure supplement 1A—see asterisks above individual bars).

Importantly, all of the synonymous variants displayed similarly neutral behavior in the pairwise and DMS competitions (Figure 6B—gray; Figure 6–figure supplement 1C), indicating that the opposing nature of these selection pressures is unlikely to reflect experimental noise.
Validation of the opposing effects of temperature and host ER proteostasis mechanisms on HA variant growth in the pairwise competitions prompted us to further examine the properties of HA sites most impacted by these selection environments. For HA variants present in each replicate mutant library (Figure 3C—Mutant libraries 1–3), we summed the mutation differential selection values at each site to calculate the net site differential selection (net site diffsel), a measure of site mutational tolerance, and averaged the net site diffsel values across the triplicate libraries (Dataset 1). Consistent with the mutation diffsel values in Figure 5A, the net site diffsel values were negatively skewed upon increased temperature, indicating reduced mutational tolerance, but positively skewed upon induction of XBP1s, indicating enhanced mutational tolerance (Figure 7A). Next, we sorted sites based on secondary structure to determine whether sites most affected by temperature or XBP1s induction belonged to particular secondary structural elements. We found that sites distributed throughout all secondary structural elements had enhanced mutational tolerance upon XBP1s induction, and that sites in strands and loops had particularly reduced mutational tolerance upon increased temperature. Thus, these selection pressures broadly impact HA mutational tolerance (Figure 7A; secondary structure).

We next plotted the net site diffsel values for sites essential for HA function and structure, to determine if these sites were impacted by temperature or ER proteostasis factors. Receptor binding sites had slightly, but significantly, enhanced mutational tolerance upon XBP1s induction, and were unaffected by increased temperature. In contrast, the mutational tolerance at antigenic sites was substantially impacted by both XBP1s induction and increased temperature (Figure 7A; functional/structural role). Moreover, sites in glycosylation sequons and cysteine residues involved in disulfide bonding, which engage glycoprotein chaperones and protein disulfide isomerases, respectively (21, 22, 28), had enhanced mutational tolerance upon XBP1s induction and reduced mutational tolerance upon increased temperature (Figure 7A; glyc. sequon/disulfide bond). Some of these sites—discussed in more detail below—were among the most affected by temperature and ER proteostasis factors.
Finally, we assessed whether ER proteostasis mechanisms and temperature differentially impact HA mutational tolerance at sites based on evolutionary conservation and surface accessibility. Specifically, we evaluated site conservation across 989 human H1 sequences from the Influenza Research Database (39). We considered completely conserved sites as conserved and all other sites as variable, and plotted the net site diffsel values for conserved and variable sites (Figure 7A; conservation), as well as the site conservation as a function of the net site diffsel (Figure 7B). This analysis illustrated that temperature and XBP1s induction affected mutational tolerance at both variable and conserved sites (Figure 7A; conservation). However, the mutational tolerance at variable sites was more affected than at conserved sites, as indicated by the net site diffsel values for conserved sites clustering near zero, whereas the net site diffsel values for variable sites were more evenly distributed along the y-axis (Figure 7B). Still, the modest impact on conserved sites suggested that host temperature and proteostasis mechanisms could modulate mutational tolerance at essentially any site in HA, regardless of the operating structural and functional constraints. This conclusion is consistent with the distribution of net site diffsel values as a function of surface accessibility, which revealed that mutational tolerance was impacted by XBP1s induction and temperature at both buried and surface-exposed sites (Figure 7A; surface accessibility), though surface-exposed sites were often more affected (Figure 7C). We note that, similar to the mutation diffsel distributions (Figure 5A), the site diffsel distributions for ATF6f/XBP1s induction mirrored those for XBP1s induction (Figure 7–figure supplement 1).

Finally, we examined the structural location of HA sites most impacted by host proteostasis mechanisms and temperature. Mapping net site diffsel values onto the HA crystal structure revealed that these selection pressures affect the mutational tolerance of sites across HA (Figure 8; Figure 8–figure supplement 1 (ATF6f/XBP1s)). Many HA sites strongly impacted by host ER proteostasis factors and temperature were in the globular head domain (Figure 8A). For example, K170, P175, and K176 were among the most mutationally intolerant sites upon
increased temperature and the most mutationally tolerant sites upon induction of XBP1s or ATF6f/XBP1s (labeled in Figure 7A and Figure 7–figure supplement 1A). These sites are on the surface of the head domain (Figure 8A), in a region where positively charged residues, such as K170 and K176, increase host receptor binding and hinder antibody escape (40). Additionally, the P175 site was the site most affected by XBP1s and ATF6f/XBP1s activation (Figure 7A and Figure 7–figure supplement 1A). Although P175 is a trans-proline in an unstructured coil, and is therefore less impactful than if found in a region of defined HA secondary structure, its mutational tolerance upon XBP1s induction could still be mediated by prolyl isomerase activity regulated by the UPR (11). Furthermore, several sites in the stem domain were strongly impacted by host ER proteostasis factors and temperature. For instance, mutational tolerance at the N497 and T499 sites was considerably reduced upon XBP1s induction, ATF6f/XBP1s induction, and upon increased temperature (Figure 8 and Figure 8–figure supplement 1), likely because these residues constitute an N-linked glycosylation sequon. N-linked glycosylation directs the co- and post-translational folding pathways of HA. Thus, removal of this glycosylation sequon may reduce the interaction of HA with the glycoprotein chaperones calnexin and calreticulin, thereby hindering HA folding and trafficking (22, 28). Altogether, host ER proteostasis factors and temperature impact the mutational tolerance at sites across the entire HA structure, irrespective of secondary structure and surface accessibility. Sites that are especially affected contain wild-type residues that likely mediate interactions between HA and host ER proteostasis factors.
**DISCUSSION**

Our results demonstrate that host proteostasis capacity and host cell temperature critically impact mutational tolerance across the entire HA protein. Activation of the XBP1s arm of the UPR globally enhanced the mutational tolerance of HA, while increased host cell temperature globally reduced the mutational tolerance of HA. Notably, the variants that were most enriched upon XBP1s induction corresponded to those most depleted upon increased temperature. This anti-correlation indicates that temperature-sensitive variants, which are likely biophysically deleterious relative to wild-type HA, are most likely to be rescued by the ER proteostasis machinery.

These selection pressures exerted similar fitness effects across the entire HA protein, influencing sites that differ in their secondary structure, surface accessibility, conservation in natural H1 sequences, and functional and structural roles. This impact is in stark contrast to selection pressures such as immune escape or drug resistance, which typically select for variants at a few specific sites. Universally modulating HA mutational tolerance suggests that host temperature and ER proteostasis mechanisms can shift the distribution of fitness effects of mutations on HA, and therefore are likely to broadly influence the accessibility of HA evolutionary trajectories. Because we observe a particularly strong impact at antigenic sites in the HA head domain (37), inhibition of ER proteostasis mechanisms may limit the accessibility of antibody escape variants. This therapeutic strategy is especially appealing because prior work has demonstrated that influenza infection can induce the unfolded protein response and that ER proteostasis mechanisms are essential for influenza propagation (26, 41).

Similar to previous studies on the role of HSP90 in the evolution of endogenous protein clients (16), our observations may derive from direct interactions between host proteostasis components and HA, or as an indirect consequence of perturbing proteostasis. The negative correlation we observe between upregulating ER proteostasis factors and increasing temperature evidences the former, because variants with compromised fitness in a biophysically challenging environment are most impacted by the ER proteostasis machinery. Alternatively, ER
proteostasis factors and increased temperature may affect the kinetics of infection or expression of third-party mediators that regulate HA function or fitness.

Irrespective of the precise origin, this study provides the first experimental evidence that ER proteostasis mechanisms define the mutational tolerance of a secretory pathway client protein, and likely extends far beyond HA and viral evolution. Prior to this work, chaperones known to potentiate and buffer protein evolution were limited to cytosolic isoforms (20), largely of HSP90 (14-19, 42). Here, we show that this role extends to ER chaperones and quality control machinery involved in the folding and assembly of membrane and secretory proteins. Understanding the evolutionary constraints of membrane and secretory proteins is critical, as they comprise highly significant therapeutic targets, including collagen (43), G-protein coupled receptors (44), antibodies (9), and viral membrane proteins (27). This work suggests that ER proteostasis mechanisms could be tuned to ameliorate protein-misfolding diseases or prevent the development of antiviral resistance. Further investigation into the impact of ER proteostasis factors on the mutational tolerance of additional secretory pathway client proteins, as well as non-client proteins, will elucidate the underlying molecular details and the pervasiveness of our findings.
MATERIALS AND METHODS

Key Resources Table

<table>
<thead>
<tr>
<th>Reagent type (species) or resource</th>
<th>Designation</th>
<th>Source or reference</th>
<th>Additional information</th>
</tr>
</thead>
<tbody>
<tr>
<td>strain, strain background (Influenza A virus)</td>
<td>Influenza A/WSN/1933</td>
<td>PMID 27271655</td>
<td></td>
</tr>
<tr>
<td>genetic reagent (Influenza A virus)</td>
<td>Influenza A/WSN/1933 Mutant Library 1</td>
<td>PMID 27271655</td>
<td></td>
</tr>
<tr>
<td>genetic reagent (Influenza A virus)</td>
<td>Influenza A/WSN/1933 Mutant Library 2</td>
<td>PMID 27271655</td>
<td></td>
</tr>
<tr>
<td>genetic reagent (Influenza A virus)</td>
<td>Influenza A/WSN/1933 Mutant Library 3</td>
<td>PMID 27271655</td>
<td></td>
</tr>
<tr>
<td>cell line (Homo sapiens)</td>
<td>HEK293\textsuperscript{ATF6f/XBP1s}</td>
<td>PMID 23583182</td>
<td></td>
</tr>
<tr>
<td>recombinant DNA reagent</td>
<td>pHW181-PB2</td>
<td>PMID 10801978</td>
<td>plasmid</td>
</tr>
<tr>
<td>recombinant DNA reagent</td>
<td>pHW182-PB1</td>
<td>PMID 10801978</td>
<td>plasmid</td>
</tr>
<tr>
<td>recombinant DNA reagent</td>
<td>pHW183-PA</td>
<td>PMID 10801978</td>
<td>plasmid</td>
</tr>
<tr>
<td>recombinant DNA reagent</td>
<td>pHW184-HA</td>
<td>PMID 10801978</td>
<td>plasmid</td>
</tr>
<tr>
<td>recombinant DNA reagent</td>
<td>pHW185-NP</td>
<td>PMID 10801978</td>
<td>plasmid</td>
</tr>
<tr>
<td>recombinant DNA reagent</td>
<td>pHW186-NA</td>
<td>PMID 10801978</td>
<td>plasmid</td>
</tr>
<tr>
<td>recombinant DNA reagent</td>
<td>pHW187-M</td>
<td>PMID 10801978</td>
<td>plasmid</td>
</tr>
<tr>
<td>recombinant DNA reagent</td>
<td>pHW188-NS1</td>
<td>PMID 10801978</td>
<td>plasmid</td>
</tr>
<tr>
<td>software, algorithm</td>
<td>dms\textunderscore tools</td>
<td>PMID 25990960</td>
<td></td>
</tr>
</tbody>
</table>

Plasmids. The following plasmids were used for generating A/WSN/1933 influenza virus:


Cell Culture. HEK293\textsuperscript{ATF6f/XBP1s} cells were cultured at 37 °C in a 5% CO\textsubscript{2} atmosphere in DMEM (CellGro) supplemented with 10% fetal bovine serum (CellGro) and 1% penicillin/streptomycin/glutamine (CellGro). Cells were generated from parental HEK293 cells (ATCC
CRL-1573; authenticated by STR-profiling) as previously described (11). Here, 1 μM TMP and 0.1 μg/mL doxycycline were used to activate ATF6 and XBP1s, respectively. All cell lines were periodically tested for mycoplasma using the MycoSensor PCR Assay Kit from Agilent (302109).

Influenza virus. All experiments were performed with influenza A/WSN/1933 (H1N1) HA viral mutant libraries (37). All infections were performed in WSN media (OptiMEM-I; Thermo Fisher Scientific) supplemented with 0.5% heat-inactivated FBS, 0.3% BSA (Invitrogen), 100 U/mL of penicillin and 100 μg/mL of streptomycin (Bio Whittaker), and 100 μg/mL of CaCl₂; swapping the inoculum with fresh WSN media 2 hours post-infection.

qPCR. HEK293<sup>ATF6/XBP1s</sup> cells were seeded at 750,000 cells/well in a 12-well plate and treated with 0.01% DMSO, 0.1 μg/mL doxycycline (Sigma), or 0.1 μg/mL doxycycline and 1 μM TMP (Alfa Aesar) for 24 h, or 100 μM arsenite for 2 h as a positive control for heat shock response activation or 10 μg/mL tunicamycin for 6 h as a positive control for unfolded protein response activation. To monitor chaperone levels during influenza infection, HEK293<sup>ATF6/XBP1s</sup> cells were infected with influenza A/WSN/1933 at an MOI of 1 for 8 h to mimic the environment of a cell infected with a single influenza virion as in the DMS. Cellular RNA was harvested using the Omega RNA Extraction kit with Homogenizer Columns. 1 μg RNA was used to prepare cDNA using random primers (total reaction volume = 20 μL; Applied Biosystems High-Capacity Reverse Transcription kit). The reverse transcription reaction was diluted to 80 μL with water, and 2 μL of each sample was used for qPCR with 2× Sybr Green (Roche) and primers for human RPLP2 (housekeeping gene), HSP70, HSP40, HSP90, ERDJ4, GRP94, BIP, SEC24D, and influenza Matrix (primer sequences in Supplementary File 1). All gene transcript levels were normalized to that of RPLP2, and the fold-change in expression relative to DMSO-treated, mock-infected cells was calculated (Figure 2–figure supplement 3C; Figure 2–source data
For qPCR of influenza-infected cells, a standard curve was prepared with a pDZ plasmid backbone containing the Influenza PR8 M segment to determine influenza Matrix copy number, which was used as a positive control for productive infection (Figure 2-source data 3).

RNA-seq. For the HEK293<sup>ATF6/XBP1s</sup> cell line characterization, HEK293<sup>ATF6/XBP1s</sup> cells were seeded at 750,000 cells/well in a 12-well plate and treated with 0.01% DMSO, 0.1 µg/mL doxycycline, or 0.1 µg/mL doxycycline and 1 µM TMP for 24 h, at either 37 °C or 39 °C. Cellular RNA was harvested using Qiagen RNeasy Plus Mini Kit with QIAshredder homogenization columns. RNA-Seq libraries were prepared using the Kapa mRNA HyperPrep RNA-seq library construction kit system and were sequenced on an Illumina HiSeq SE40.

RNA-seq Analysis (46). Quality control: Reads were aligned against hg19 (Feb., 2009) using bwa mem v. 0.7.12-r1039 [RRID:SCR_010910] with flags –t 16 –f and mapping rates, fraction of multiply-mapping reads, number of unique 20-mers at the 5´ end of the reads, insert size distributions and fraction of ribosomal RNAs were calculated using bedtools v. 2.25.0 [RRID:SCR_006646] (47). In addition, each resulting bam file was randomly down-sampled to a million reads, which were aligned against hg19 and read density across genomic features were estimated for RNA-Seq-specific quality control metrics.

RNA-Seq mapping and quantitation: Reads were aligned against GRCh38 / ENSEMBL 89 annotation using STAR v. 2.5.3a with the following flags -runThreadN 8 --runMode alignReads --outFilterType BySJout --outFilterMultimapNmax 20 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --outFilterMismatchNmax 999 --alignIntronMin 10 --alignIntronMax 1000000 --alignMatesGapMax 1000000 --outSAMtype BAM SortedByCoordinate --quantMode TranscriptomeSAM with --genomeDir pointing to a 75nt-junction GRCh38 STAR suffix array(48). Gene expression was quantitated using RSEM v. 1.3.0 [RRID:SCR_013027] with the following
flags for all libraries: rsem-calculate-expression --calc-pme --alignments -p 8 --forward-prob 0 against an annotation matching the STAR SA reference (49). Posterior mean estimates (pme) of counts and estimated RPKM were retrieved.

Differential expression analysis: Treatments were compared against DMSO-treatment (Basal) either at 37 °C or 39 °C for each condition (Figure 2; Figure 2–figure supplement 1; Figure 2–source data 1). Briefly, differential expression was performed in the R statistical environment (R v. 3.4.0) using Bioconductor’s DESeq 2 package on the protein-coding genes only [RRID:SCR_000154] (50). Dataset parameters were estimated using the estimateSizeFactors(), and estimateDispersions() functions; read counts across conditions were modeled based on a negative binomial distribution, and a Wald test was used to test for differential expression (nbinomWaldtest(), all packaged into the DESeq() function), using the treatment type as a contrast. Fold-changes and p-values were reported for each protein-coding gene. Heat maps were generated in Spotfire (Tibco), using a complete linkage function and the cosine correlation as distance metrics (Figure 2).

Comparison of upregulated genes in present study to previous study (11): This analysis was performed to compare genes upregulated upon XBP1s and ATF6f/XBP1s induction at 37 °C, relative to DMSO treatment at 37 °C. The analysis was restricted to genes with clear representation in both the whole-genome array (11) and the RNA-Seq annotation (n = 13058). Transcripts detected by the two methods were matched based on an Affymetrix probe ID for the whole-genome array and an ENSEMBL ID based on DAVID v6.8 for the RNA-Seq (46). Residual unmatched transcripts were matched based on NCBI IDs, and finally based on gene names (correcting for updates in official gene symbols between annotations). Both datasets were filtered applying a 2-fold change cutoff and an FDR of 0.05. The resulting overlap is presented in Figure 2–figure supplement 2. The genes for each subset of the Venn Diagram (Figure 2–figure supplement 2) are listed in Figure 2–source data 2. For each subgroup, ENSEMBL identities were retrieved for each gene and used in the DAVID online Gene Ontology environ-
ment (46) to call enriched functional categories against the background of 13058 genes jointly
called across both RNA quantification methods (Figure 2–source data 2).

Cell growth assay. HEK293^{ATF6/XBP1s} cells were seeded at 750,000 cells/well in poly-D-lysine
(Sigma)-coated (0.05 mg/mL in PBS, 37 °C for 15 min) 12-well plates and pre-treated with
0.01% DMSO, 0.1 µg/mL doxycycline, or 0.1 µg/mL doxycycline and 1 µM TMP, and placed at
either 37 °C or 39 °C. To mimic the infection conditions, the cellular growth media was replaced
with WSN media supplemented with 0.01% DMSO, 0.1 µg/mL doxycycline, or 0.1 µg/mL
doxycycline and 1 µM TMP 18 h post-seeding, returning plates to either 37 °C or 39 °C. 48 h
after the mock infection, the media was replaced with WSN media containing 50 µM resazurin
sodium salt (Sigma). After 4 h of incubation, 100 µL of media was used to quantify resorufin flu-
orescence (excitation 530 nm; emission 590 nm) using a Take-3 plate reader (BioTeK) (Figure
2–figure supplement 3A).

Influenza infection in modified host cell environments. HEK293^{ATF6/XBP1s} cells were seeded
at 750,000 cells/well in poly-D-lysine-coated 12-well plates and treated with 0.01% DMSO, 0.1
µg/mL doxycycline, or 0.1 µg/mL doxycycline and 1 µM TMP, at either 37 °C or 39 °C, for 18 h.
To assess viral growth in each selection condition, pre-treated cells were infected with influenza
A/WSN/1933 at an MOI of 0.01 virions/cell for 48 h, at either 37 °C or 39 °C, to mimic DMS and
pairwise competition conditions. Viral supernatant was harvested 48 hours post-infection and
titered using a TCID_{50} assay (Figure 2–figure supplement 3B).

Deep mutational scanning. Three independently generated HA mutant plasmid libraries were
used to generate three biological replicate viral libraries in Doud et al (37); these viral libraries
were generously provided by Jesse Bloom (Fred Hutchinson Cancer Research Center).
HEK293ATF6iXBP1s cells were plated in poly-D-lysine-coated 15 cm plates at a density of $50 \times 10^6$ cells per dish; 8 hours after plating, cells were treated with 0.01% DMSO, 0.1 $\mu$g/mL doxycycline, or 0.1 $\mu$g/mL doxycycline and 1 $\mu$M TMP for 16 h, at either 37 °C or 39 °C. 16 h after treatment, $1 \times 10^6$ Infectious virions (determined using a TCID$_{50}$ assay) from each viral library were used to infect two 15 cm plates in each condition at an MOI of 0.01 virions/cell. In addition, one 15 cm plate at both 37 °C and 39 °C was either mock infected (negative control) or infected with wild-type virus. For infection, the cellular growth media was replaced with WSN media containing a mutant virus library, wild-type virus, or no virus for mock infection. After 2 h, the inoculum was replaced with fresh WSN media containing 0.01% DMSO, 0.1 $\mu$g/mL doxycycline, or 0.1 $\mu$g/mL doxycycline and 1 $\mu$M TMP. 48 hours post-infection, the viral supernatant was harvested, centrifuged at 1000 $\times$ g for 5 min to remove cell debris, and stored at –80 °C. To extract viral RNA from the viral supernatant, the supernatant was thawed and virions were concentrated by ultracentrifugation, spinning at 25,000 rpm for 2 h at 4 °C (SW 32 Ti swinging bucket rotor, Beckman Coulter). The supernatant was then decanted, and the virion pellet was resuspended in 140 $\mu$L QIAgen Viral RNA Buffer AVL. Viral RNA was then extracted using a QIAgen Viral RNA mini kit, per the manufacturer’s instructions, changing collection tubes at each step. Viral RNA was reverse transcribed using the AccuScript High Fidelity 1st strand cDNA synthesis kit (Agilent) using 5'-WSN-HA and 3'-WSN-HA primers (primer sequences in Supplementary File 1). At least $10^6$ HA molecules were PCR-amplified for preparation of subamplicon sequencing libraries, as previously described (37), to ensure sufficient sampling of viral library diversity. Briefly, this sequencing library preparation method appends unique, random barcodes and part of the Illumina adapter to HA subamplicon molecules. In a second round of PCR, the complexity of the uniquely barcoded subamplicons was controlled to be less than the sequencing depth, and the remainder of the Illumina adapter is appended (37). The resulting libraries were sequenced on an Illumina HiSeq 2500 in rapid run mode with 2 × 250 bp paired-end reads.
Deep mutational scanning data analysis. dms_tools (http://jbloomlab.github.io/dms_tools/) (38) was used to align reads to the Influenza A/WSN/1933 HA reference sequence, count amino acid variants across HA, and calculate the differential selection for each variant between two selection conditions, as previously described (23). Briefly, reads were trimmed, aligned to the Influenza A/WSN/1933 reference sequence, and read pairs were quality filtered by discarding read pairs with low-quality sites in the barcode region of either read, in addition to read pairs with more than 7.5% low quality sites. The remaining read pairs were sorted by barcode and barcodes with less than 2 read pairs were discarded. For each remaining barcode, mutation calls were only made when greater than 75% of reads concurred (37, 38). For each sample, the numbers of read pairs, barcoded read pairs, and reads per barcode are reported in Dataset 1 (see sample _summarystats.txt) and the sequencing data analyses are available at https://github.com/amphilli/HA_DMS_2018. For evaluating mutational differential selection, variants not present in each replicate starting library were removed from the analysis and were subsequently filtered for variants that incurred selection of the same sign across biological triplicates (Figure 5–source data 1). The mutation differential selection values for these filtered variants were visualized on sequence logo plots, made using dms_tools (Figure 4; Figure 4–figure supplements 1–5) (23, 38). Correlation between mutation differential selection values across selection conditions was determined using linear regression using Graph Pad Prism software, reporting Pearson correlation coefficients (Figures 5B–C; Figure 5–figure supplement 1). For each site in HA, the net site differential selection (net site diffsel) was calculated by summing the mutation differential selection values separately for each replicate starting library, excluding mutational differential selection values for variants that were not present in the respective starting library. These replicate net site diffsel values were then then averaged and sorted by HA secondary structure, relative surface accessibility (32, 51), receptor binding and antigenic sites (37), glycosylation sequons and disulfide bond sites (21, 22, 28), and
natural HA site conservation (39) (Figure 7; Figure 7–figure supplement 1; Figure 7–source data 1). HA secondary structure and relative surface accessibility were determined using DSSP (51, 52) based on the HA crystal structure (PDBID 1RVX) (32). In Figure 7A, sites designated buried were the 50% least surface-accessible sites; sites designated exposed were the 50% most surface-accessible sites. Receptor binding and antigenic sites were classified as in Doud et al, and included sites nearby antigenic sites (37). Natural HA site conservation was based on 989 non-redundant human H1 sequences preceding the 2009 H1N1 pandemic, accessed from the NIAID Influenza Research Database (accessed Jan. 10th, 2018) (39). Site entropies were calculating using the Influenza Research Database SNP Analysis Tool (39), and were normalized to a scale of 0–100, with a site entropy of 100 representing the most variable site in HA and 0 representing the most conserved site (Figure 7B). In Figure 7A, sites designated conserved had a site entropy of zero; all other sites were designated as variable. The net site differential selection values were mapped onto the HA crystal structure (PDBID 1RVX) (32) using PyMOL (53) (Figure 8; Figure 8–figure supplement 1).

Data availability. FASTQ files for DMS sequencing are available in the Sequence Read Archive under accession number SRP149672. The deep mutational scanning data analyses are available at https://github.com/amphilli/HA_DMS_2018, and is also available in Dataset 1. All mutation differential selection values from deep mutational scanning (pre- and post-filtering) are available in Figure 5–source data 1. The complete RNAseq data are available from GEO under accession number GSE115168.

Infectious viral titering via tissue culture infectious dose (TCID$_{50}$) assay. 10-fold dilutions of each virus were prepared in quadruplicate in 96-well plates. 5,000 MDCK-SIAT1 cells were then added to each well and incubated at 37 °C for 72 h, after which the wells were scored for the presence of cytopathic effect. The dilutions of virus displaying cytopathic effect in the MDCK-
SIAT1 cells were then used to calculate the TCID\textsubscript{50}/μL using https://github.com/jbloomlab/reedmuenchcalculator as described by Thyagarajan and Bloom (54), where virions/μL = 0.69*TCID\textsubscript{50}/μL.

Reverse genetics pairwise viral competitions. HA mutant-encoding plasmids were prepared by introducing point mutations into the pHW184-HA plasmid using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent 200522). Mutant viruses were generated from plasmids by transfecting a co-culture of 2.5 x 10\textsuperscript{4} MDCK-SIAT1 and 3 x 10\textsuperscript{5} HEK 293T cells, as previously described (45), followed by titering using the TCID\textsubscript{50} assay. For each competition, HEK293\textsuperscript{ATF6\textsuperscript{f}/XBP1\textsuperscript{s}} cells were plated in poly-D-lysine-coated 24-well dishes at 3.5 x 10\textsuperscript{5} cells/well and treated 8 h later with 0.01% DMSO, 0.1 μg/mL doxycycline, or 0.1 μg/mL doxycycline and 1 μM TMP at either 37 °C or 39 °C. 24 h after plating, cells were infected with a 1:1 mixture of wild-type and mutant viruses at an MOI of 0.01 virions/cell in biological triplicate under conditions identical to that of the deep mutational scanning experiment. After 2 h, the inoculum was replaced with fresh WSN media supplied with 0.01% DMSO, 0.1 μg/mL doxycycline, or 0.1 μg/mL doxycycline and 1 μM TMP. 48 h post-infection, infectious supernatant was harvested, centrifuged at 1000 × g for 5 min to remove cell debris, and stored at –80 °C. Viral RNA was extracted from 140 μL infectious supernatant using the QIAamp Viral RNA Mini kit and at least 10\textsuperscript{6} HA molecules were reverse transcribed using SuperScript III Reverse Transcriptase (Thermo Fisher Scientific) with 5'-WSN-HA and 3'-WSN-HA primers (Supplementary File 1). The dsDNA was purified twice using 0.9× AMPure XP beads (Beckman Coulter) and quantified using a Quant-iT PicoGreen assay (Life Technologies). Illumina NexteraXT sequencing libraries were prepared using a Mosquito HTS liquid handler (TTP Labtech) and sequenced on an Illumina HiSeq with 50-bp single-end reads.
**Pairwise competition sequencing data analysis.** To overcome biases in read mapping from consecutive, multi-nucleotide sequence variants, mapping target sequences were generated for each Influenza A/WSN/1933 HA variant. Briefly, regions flanking the variant (given the experimental read length) would be identified, relevant groups of bases substituted with their engineered variants, and the rest of the sequence substituted with “N” in order to maintain the sequence position of said variants, thus generating sets of ‘pseudo-haplotypes’. Sequencing reads were aligned to the wild-type Influenza A/WSN/1933 HA sequence and the set of pseudo-haplotypes using bwa mem (v. 0.7.12-r1039) (arXiv:1303.3997), with flag –t 16, and sorted and indexed bam files were generated using samtools (v 1.3) (55). These bam files were processed using samtools mpileup with flags –excl-flags 2052, -d 30000000 and the same reference sequences used for mapping (56). Allele counts at each position of the wild-type and pseudo-haplotype sequences were summed and reported along the wild-type Influenza A/WSN/1933 HA sequence for each sample for calculation of variant frequencies. The change in variant frequency upon selection was normalized to that of wild-type, thus providing diffsel values on the same scale as the DMS competition (**Figure 6B; Figure 6–figure supplement 1**).

**Statistics.** All experiments were performed in biological triplicate with replicates defined as independent experimental entireties (i.e., from plating the cells to acquiring the data). For deep mutational scanning, each biological replicate mutant viral library was prepared from independently generated mutant plasmid libraries, as previously reported (37). Differential selection values from deep mutational scanning, as well as selection values from pairwise competitions, were tested for significance of deviation from zero (wild-type behavior), using a one-sample $t$-test in Graph Pad Prism (**Figure 5A; Figure 6–figure supplement 1**). Differential selection values from pairwise competitions were compared between selection conditions by a Student’s $t$-test (**Figure 6–figure supplement 1**). Correlation between diffsel values was determined by
linear regression using Graph Pad Prism software, reporting Pearson correlation coefficients and $p$-values from $F$-tests, testing significance of slope deviating from zero (Figures 5B–C; Figure 5–figure supplement 1). Correlation between site diffsel values ($N = 565$) was determined by linear regression using Pandas Python package, reporting Pearson correlation coefficients (Figure 3–source data 1). For DMS mutational tolerance analyses, statistical significance of the deviation of the mean of the net site diffsel distribution from zero (wild-type behavior) was tested by a Student’s $t$-test; one-tailed $p$-values are reported, to assess whether the net site diffsel distributions for XBP1s and ATF6f/XBP1s induction are significantly greater than zero, and whether the net site diffsel distribution for increased temperature is significantly less than zero (Figure 7A; Figure 7–figure supplement 1; Figure 7–source data 1).
ACKNOWLEDGEMENTS

The authors would like to thank Christopher E. R. Richardson (MIT), Andrew S. DiChiara (AbbVie), and Chet M. Berman (MIT) for feedback on this manuscript.

COMPETING INTERESTS

The authors declare no competing interests.
References


22. Daniels R, Kurowski B, Johnson AE, Hebert DN. N-Linked glycans direct the
23. Doud MB, Hensley SE, Bloom JD. Complete mapping of viral escape from neutralizing
influenza virus hemagglutinin to analogs of its cell-surface receptor, sialic acid: analysis by
proton nuclear magnetic resonance spectroscopy and X-ray crystallography. Biochemistry.
of the defects in the hemagglutinin gene of two temperature-sensitive mutants of A/WSN/33
virus hemagglutinin glycoproteins by the host endoplasmic reticulum (ER) stress pathway
2018;92(1):e01690-17.
27. Hurtley SM, Bole DG, Hoover-Litty H, Helenius A, Copeland CS. Interactions of
28. Hebert DN, Zhang JX, Chen W, Foellmer B, Helenius A. The number and location of
glycans on influenza hemagglutinin determine folding and association with calnexin and
29. Ueda M, Sugiura A. Physiological characterization of influenza virus temperature-
30. Skehel JJ, Wiley DC. Receptor binding and membrane fusion in virus entry: the
Stability of the influenza virus hemagglutinin protein correlates with evolutionary dynamics.
mSphere. 2018;3(1):e00554-17.
32. Gamblin SJ, Haire LF, Russell RJ, Stevens DJ, Xiao B, Ha Y, et al. The structure and
receptor binding properties of the 1918 influenza hemagglutinin. Science. 2004;303(5665):1838-
42.
2015;528(7583):510-6.
35. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of
technical reproducibility and comparison with gene expression arrays. Genome research.
2008;18(9):1509-17.
36. Russell AB, Trapnell C, Bloom JD. Extreme heterogeneity of influenza virus infection in
37. Doud MB, Bloom JD. Accurate measurement of the effects of all amino-acid mutations
38. Bloom JD. Software for the analysis and visualization of deep mutational scanning data.
Research Database: An integrated bioinformatics resource for influenza virus research. Nucleic
Acids Res. 2017;45(D1):D466-74.
Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. Science.


DeLano WL. The PyMOL molecular graphics system. 2002.


Figure 1. The proteostasis boundary model for protein evolution. (A) The accessibility of protein evolutionary intermediates is defined by protein stability, folding rate, and misfolding rate. Protein variants inside the boundary are accessible (shaded in black), and those outside the boundary are inaccessible (shaded in red). Figure generated using modified Mathematica macro from Powers et al (7). (B) In a biophysically permissive environment, such as cells with elevated levels of proteostasis factors, the boundary may be shifted, making otherwise non-functional variants accessible.

Figure 2. Induction of XBP1s and ATF6f/XBP1s is selective and does not globally induce stress responses. (A) Heat map of selected transcript-level effects of upregulating ER folding machinery in HEK293ATF6f/XBP1s cells via XBP1s or ATF6f/XBP1s at 37 °C, relative to a basal environment at 37 °C. (B) Heat map of selected transcript-level effects of upregulating ER proteostasis machinery via XBP1s or ATF6f/XBP1s at 39 °C, relative to a basal environment at 39 °C. (C) Heat map of selected transcript-level effects of increasing the temperature from 37 °C to 39 °C in a basal ER proteostasis environment for 24 h. Genes in A–C grouped by ER (top) or cytosolic (bottom) proteostasis networks, and also based on functional classifications (11). For A–B, tetracycline repressor-regulated XBP1s was induced by treatment with 0.1 μg/mL doxycycline (24 h); destabilized domain-regulated ATF6f was induced by treatment with 1 μM trimethoprim (24 h). Figure 2–figure supplement 1. Full transcriptome analysis confirmed selectivity of XBP1s and ATF6f/XBP1s induction. Figure 2–figure supplement 2. Comparison of HEKATF6f/XBP1s RNA-Seq characterization to previous whole-genome array characterization. Figure 2–figure supplement 3. Methods to induce XBP1s and ATF6f/XBP1s did not cause cytotoxicity and were functional during influenza infection. Figure 2–source data 1. Complete RNASeq differen-
tial expression analysis. Figure 2–source data 2. Comparison of HEK^{ATF6f/XBP1s} characterization to previous characterization. Figure 2–source data 3. Fold-change of HSR and UPR targets during influenza infection.

Figure 2–figure supplement 1. Full transcriptome analysis confirmed selectivity of XBP1s and ATF6f/XBP1s induction. Volcano plots show change in transcript expression upon activating XBP1s alone or with ATF6f at 37 °C (A–B) and 39 °C (C–D) in HEK293^{ATF6f/XBP1s} cells. (E) Change in transcript expression upon increasing temperature in basal cells. In A–E, transcripts displaying >2-fold change in expression with p-values < 0.00001 are plotted in red, with select outliers labeled. Figure 2–source data 1. Complete RNAseq differential expression analysis.

Figure 2–figure supplement 2. Comparison of HEK^{ATF6f/XBP1s} RNA-Seq characterization to previous whole-genome array characterization. Number of upregulated genes in current study represented in red (as determined by RNA-Seq following 24 h treatment with 0.1 μg/mL doxycycline and 1 μM trimethoprim in low serum media); number of upregulated genes observed in previous study represented in blue (as determined by whole-genome array following 12 h treatment with 1 μg/mL doxycycline and 10 μM trimethoprim in complete media (11)). For this comparison, genes were considered upregulated if they increased in expression by more than 2-fold, using a FDR of 0.05. This filtering is more stringent than that reported in Shoulders et al (11), but was necessary to compare data generated from RNA-Seq and a whole genome array. Figure 2–source data 2. Comparison of HEK^{ATF6f/XBP1s} characterization to previous characterization.

Figure 2–figure supplement 3. Methods to induce XBP1s and ATF6f/XBP1s are not cytotoxic and are functional during influenza infection. (A) Induction of XBP1s or ATF6f/XBP1s and in-
creased temperature (72 h) does not alter metabolic activity in the absence of influenza, as measured by a resazurin assay. (B) Wild-type WSN/1933 influenza titers (determined by TCID$_{50}$ assay) from infections in modulated cellular environments (MOI = 0.01; 48 hours post-infection).

For A–B, the average of biological triplicates is plotted with error bars representing the SEM. (C) Fold change of heat shock response (HSR) and unfolded protein response (UPR) gene transcripts in modulated ER proteostasis environments with and without influenza (MOI = 1, 8 hours post-infection), relative to a mock-infected basal environment at 37 °C. UPR gene transcripts are designated as ATF6f or XBP1s targets based on prior studies (11). Positive controls for activating the UPR (treatment with the N-glycosylation inhibitor tunicamycin) and the heat shock response (HSR) (treatment with arsenite to induce oxidative stress) are shown. **Figure 2–source data 3.** Fold-change of HSR and UPR targets during influenza infection.

**Figure 3.** Deep mutational scanning of HA in modulated ER proteostasis environments. (A) Scheme for performing deep mutational scanning, and (B) quantifying the differential selection (diffsel) for each mutant in the selected environments, which can be visualized on sequence logo plots. In A, XBP1s or ATF6f/XBP1s were induced at either 37 °C or 39 °C for 16 h prior to the selection, which consisted of a 48 h infection. In B, $f_{\text{WT}}$ and $f_{\text{mut}}$ denote the wild-type and mutant frequencies, respectively. (C) Percentage of amino acid space sampled by each replicate library prior to filtering, and the percentage sampled by all three replicate mutant libraries. **Figure 3–source data 1.** Correlation coefficients for unfiltered absolute site differential selection values between biological replicate selections.

**Figure 4.** HA variants were depleted at increased temperature but enriched upon XBP1s or ATF6f/XBP1s induction. Cropped sequence logo plots show differential selection on HA at positions where variants were depleted upon increased temperature but enriched upon XBP1s and ATF6f/XBP1s induction (based on sequential numbering of WSN HA). Size of amino acid ab-
breviation corresponds to magnitude of selection. Only variants behaving consistently across biological triplicates are plotted; all selections are plotted on the same scale. Figure 4–figure supplement 1. Full WSN HA sequence logo plot: Basal 39 °C vs. Basal 37 °C. Figure 4–figure supplement 2. Full WSN HA sequence logo plot: XBP1s 37 °C vs. Basal 37 °C. Figure 4–figure supplement 3. Full WSN HA sequence logo plot: XBP1s 39 °C vs. Basal 39 °C. Figure 4–figure supplement 4. Full WSN HA sequence logo plot: ATF6f/XBP1s 37 °C vs. Basal 37 °C. Figure 4–figure supplement 5. Full WSN HA sequence logo plot: ATF6f/XBP1s 39 °C vs. Basal 39 °C. Dataset 1. Complete analysis of deep mutational scanning data.

Figure 4–figure supplement 1. Full WSN HA sequence logo plot: Basal 39 °C vs. Basal 37 °C. Logo plot displays variants that behaved consistently across three biological replicates. Variants above the line (representative of selection on wild-type) were enriched at 39 °C relative to 37 °C in a basal environment; variants below the line were depleted at 39 °C. The wild-type WSN HA residue is shown below the corresponding logo, with the sites numbered below the wild-type sequence (based on sequential numbering of WSN HA). The size of the amino acid letter abbreviation is proportional to the diffsel for that amino acid variant, and all logo plots are plotted on the same scale.

Figure 4–figure supplement 2. Full WSN HA sequence logo plot: XBP1s 37 °C vs. Basal 37 °C. Logo plot displays variants that behaved consistently across three biological replicates. Variants above the line (representative of selection on wild-type) were enriched upon XBP1s induction at 37 °C; variants below the line were depleted. The wild-type WSN HA residue is shown below the corresponding logo, with the sites numbered below the wild-type sequence (based on sequential numbering of WSN HA). The size of the amino acid letter abbreviation is proportional to the diffsel for that amino acid variant, and all logo plots are plotted on the same scale.
Figure 4–figure supplement 3. Full WSN HA sequence logo plot: XBP1s 39 °C vs. Basal 39 °C. Logo plot displays variants that behaved consistently across three biological replicates. Variants above the line (representative of selection on wild-type) were enriched upon XBP1s induction at 39 °C; variants below the line were depleted. The wild-type WSN HA residue is shown below the corresponding logo, with the sites numbered below the wild-type sequence (based on sequential numbering of WSN HA). The size of the amino acid letter abbreviation is proportional to the diffsel for that amino acid variant, and all logo plots are plotted on the same scale.

Figure 4–figure supplement 4. Full WSN HA sequence logo plot: ATF6f/XBP1s 37 °C vs. Basal 37 °C. Logo plot displays variants that behaved consistently across three biological replicates. Variants above the line (representative of selection on wild-type) were enriched upon ATF6f/XBP1s induction at 37 °C; variants below the line were depleted. The wild-type WSN HA residue is shown below the corresponding logo, with the sites numbered below the wild-type sequence (based on sequential numbering of WSN HA). The size of the amino acid letter abbreviation is proportional to the diffsel for that amino acid variant, and all logo plots are plotted on the same scale.

Figure 4–figure supplement 5. Full WSN HA sequence logo plot: ATF6f/XBP1s 39 °C vs. Basal 39 °C. Logo plot displays variants that behaved consistently across three biological replicates. Variants above the line (representative of selection on wild-type) were enriched upon ATF6f/XBP1s induction at 39 °C; variants below the line were depleted. The wild-type WSN HA residue is shown below the corresponding logo, with the sites numbered below the wild-type sequence (based on sequential numbering of WSN HA). The size of the amino acid letter abbreviation is proportional to the diffsel for that amino acid variant, and all logo plots are plotted on the same scale.
Figure 5. ER proteostasis mechanisms and temperature divergently impact HA variant viral growth. (A) Average differential selection (diffsel) for each HA variant; the black line designates the mean of the distribution. Significance of deviation of the mean from zero (no differential selection) was tested by a one-sample t-test; *** designates two-tailed p-values ≤ 0.001. Variant diffsel values are staggered about the x-axis to minimize overlap. The number of variants with diffsel > 0 and < 0 is listed below each distribution. (B) Correlation for diffsel values for inducing XBP1s at 39 °C versus increasing the temperature in basal cells (N = 271). (C) Correlation for diffsel values for inducing ATF6f/XBP1s at 39 °C versus increasing the temperature in basal cells (N = 320). For A–C, average diffsel values for variants behaving consistently across biological triplicates are plotted. For B–C, Pearson correlation and p-value from an F-test are shown. Figure 5–figure supplement 1. Induction of ATF6f/XBP1s and XBP1s similarly impact HA variant viral growth. Figure 5–source data 1. Differential selection values for each selection, pre- and post-filtering.

Figure 5–figure supplement 1. Induction of ATF6f/XBP1s and XBP1s similarly impact HA variant viral growth. Correlation for diffsel values for inducing XBP1s at 39 °C versus inducing ATF6f/XBP1s at 39 °C (N = 261). Average diffsel values for variants behaving consistently across biological triplicates are plotted; Pearson correlation and p-value from an F-test are shown.

Figure 6. Divergent fitness effects revealed by DMS assessed by pairwise competitions. (A) Scheme for pairwise competition between wild-type HA and mutant HA influenza. Detailed description of competition conditions in Figure 6–figure supplement 1. (B) The difference between the average diffsel upon increased temperature (ΔT) and the average diffsel upon induction of XBP1s or ATF6f/XBP1s (ΔUPR) is plotted. This difference from the pairwise competitions

38
is plotted on the y-axis, with error bars representing SEM for biological triplicates, and the corresponding difference from the DMS is plotted on the x-axis. Variants that validated (i.e., diffsel upon increased temperature was significantly different from that upon UPR induction in both DMS and pairwise competitions) are in black; variants that did not validate are in red; synonymous (neutral) validated negative controls are in gray. All non-synonymous variants are labeled.

**Figure 6—figure supplement 1.** Individual pairwise competition diffsel values.

**Figure 6—figure supplement 1.** Individual pairwise competition diffsel values. (A) Diffsel of validated variants in pairwise competition with wild-type influenza upon induction of XBP1s or ATF6f/XBP1s and increased temperature. Variants were considered “validated” if the diffsel upon induction of XBP1s or ATF6f/XBP1s was significantly different from that upon increased temperature. (B) Diffsel of non-validated variants in pairwise competition. (C) Diffsel of synonymous mutants in pairwise competition. For A–C, the average of biological triplicates is plotted with error bars corresponding to SEM. Asterisks comparing selection conditions represent significance evaluated by Student’s t-test, while asterisks comparing an individual selection condition to no selection (diffsel of zero) represent significance evaluated by a one-sample t-test. For all significance tests, *, **, and *** represent two-tailed p-values ≤ 0.05, 0.01, and 0.001, respectively; “ns” corresponds to a p-value > 0.05. Induction of UPR conditions in green correspond to ∆UPR in **Figure 6**; increased temperature conditions in blue correspond to ∆T in **Figure 6**.

**Figure 7.** XBP1s induction and temperature impacted HA mutational tolerance at sites across HA with diverse functional and structural roles. (A) Average net site diffsel (sum of mutation diffsel values at a given site) is plotted for increased temperature in a basal environment (blue) and upon induction of XBP1s at 39 °C (purple). Sites are sorted by secondary structure, functional/structural role, conservation in H1 sequences, and surface accessibility (buried residues are the 50% least surface-accessible sites). Select outliers are labeled. Means are represented...
by black lines and significance of deviation from zero (no selection) was determined by a one-sample \( t \)-test and is indicated above each distribution; * ,**, and *** represent one-tailed \( p \)-values \( \leq 0.05 \), 0.01, and 0.001, respectively; ns corresponds to a \( p \)-value > 0.05. (B) Average net site diffsel is plotted as a function of site conservation, where \( x = 0 \) corresponds to a completely conserved site among 989 non-redundant human H1 sequences (1918–2008) (39). (C) Average net site diffsel is plotted as a function of site surface accessibility, where \( x = 0 \) corresponds to a buried residue in the (PDB ID 1RVX (32)). Figure 7–figure supplement 1. ATF6f/XBP1s induction and temperature impacted HA mutational tolerance at sites across HA with diverse functional and structural roles. Figure 7–source data 1. Properties and differential selection for each site of HA.
Figure 8. Mutational tolerance at sites across the mature HA protein were impacted by XBP1s induction and increased temperature. (A) Average net site diffsel values (which correspond to mutational tolerance) are mapped onto the HA crystal structure PDBID 1RVX (32), with select outliers from Figure 7 labeled. (B) Cartoon representation of A displaying secondary structural elements. In A–B, negative net site diffsel values are mapped in blue; positive net site diffsel values are mapped in red. Figure 8–figure supplement 1. Mutational tolerance at sites across the mature HA protein were impacted by ATF6f/XBP1s induction and increased temperature.

Figure 8–figure supplement 1. Mutational tolerance at sites across the mature HA protein were impacted by ATF6f/XBP1s induction and increased temperature. (A) Average net site diffsel values (which correspond to mutational tolerance) are mapped onto the HA crystal structure PDBID 1RVX (32), with select outliers from Figure 7–figure supplement 1 labeled. (B) Cartoon representation of A displaying secondary structural elements. In A–B, negative net site diffsel values are mapped in blue; positive net site diffsel values are mapped in red.
Supplementary File 1. Primer sequences for qPCR and HA sequencing.

Dataset 1. Complete analysis of deep mutational scanning data. This folder is a compressed version of the Github repository available at https://github.com/amphilli/HA_DMS_2018.

Figure 2–source data 1. Complete RNAseq differential expression analysis.

Figure 2–source data 2. Comparison of HEK\textsuperscript{ATF6f/XBP1s} characterization to previous characterization. Genes are tabulated as upregulated in the current study (RNAseq), the previous study (Array) alone (11), or upregulated in both studies, as illustrated in Figure 2–figure supplement 2. For each gene subset, the corresponding GO group enrichment analysis is provided.

Figure 2–source data 3. Fold change of HSR and UPR targets during influenza infection.

Figure 3–source data 1. Correlation coefficients for unfiltered absolute site differential selection values between biological replicate selections. N = 565.

Figure 5–source data 1. Differential selection values for each selection, pre- and post-filtering.

Figure 7–source data 1. Properties and differential selection for each site of HA.
A Biophysically permissive environment (i.e. elevated levels of proteostasis factors) creates a more stable environment for proteins. In A, the basal environment has a slow folding rate and a high misfolding rate, resulting in unstable protein. In B, the biophysically permissive environment has a fast folding rate and a low misfolding rate, allowing protein inside the expanded boundary to function.
A) XBP1s 37 °C vs. Basal 37 °C

- log10 (p-value)

- log2 fold change over basal 37 °C

B) ATF6f/XBP1s 37 °C vs. Basal 37 °C

- log10 (p-value)

- log2 fold change over basal 37 °C

C) XBP1s 39 °C vs. Basal 39 °C

- log10 (p-value)

- log2 fold change over basal 39 °C

D) ATF6f/XBP1s 39 °C vs. Basal 39 °C

- log10 (p-value)

- log2 fold change over basal 39 °C

E) Basal 39 °C vs. Basal 37 °C

- log10 (p-value)

- log2 fold change over basal 37 °C

- log2 fold change over basal 39 °C
Comparison of number of upregulated genes to previous characterization

XBP1s 37 °C vs. Basal 37 °C
- 201 genes
- 65 genes

ATF6f/XBP1s 37 °C vs. Basal 37 °C
- 268 genes
- 120 genes
- 3 genes

Key:
- eLife 2018: RNA-Seq
  - 0.5% FBS
  - 24 h treatment
  - 0.1 μg/mL doxycycline
  - + 1 μM trimethoprim (ATF6f)

- Cell Rep., 2013: Whole-genome array
  - 10% FBS
  - 12 h treatment
  - 1 μg/mL doxycycline (XBP1s)
  - + 10 μM trimethoprim (ATF6f)
HSP90
HSP70
HSP40
GRP94
BiP
ERDJ4
SEC24D

Control: UPR activation
Control: HSR activation
Basal 37 °C
+XBP1s 37 °C
+ATF6f/XBP1s 37 °C
Basal 39 °C
+XBP1s 39 °C
+ATF6f/XBP1s 39 °C

Key for A,B:
- Basal 37 °C
- Basal 39 °C
- +XBP1s 37 °C
- +XBP1s 39 °C

Viability of mock-infected cells

Cellular metabolic activity (fluorescence units)

Viral titers for infections in modulated cellular proteostasis environments

Viral titer (TCID₅₀/μL)

Fold change relative to mock-infected basal, 37 °C:
HA viral mutant library pre-selection

37 °C

+XBP1s
+ATF6f/XBP1s

39 °C

+XBP1s
+ATF6f/XBP1s

HA viral mutant library post-selection

sequence to quantify differential selection (diffsel)

diffsel_{WT} = 0

WT sequence site

M K A K L

enriched in selection
debated in selection

diffsel = log_2(f_{mut}/f_{WT}(selection))

(f_{mut}/f_{WT}(mock-selection))

Percentage amino acid space sampled

Mutant library 1 44.8%
Mutant library 2 50.2%
Mutant library 3 42.7%
Triplicate mutant libraries 34.1%
<table>
<thead>
<tr>
<th></th>
<th>Basal 39 °C vs. Basal, 37 °C</th>
<th>XBP1s 37 °C vs. Basal, 37 °C</th>
<th>XBP1s 39 °C vs. Basal, 39 °C</th>
<th>ATF6f/XBP1s 37 °C vs. Basal, 37 °C</th>
<th>ATF6f/XBP1s 39 °C vs. Basal, 39 °C</th>
</tr>
</thead>
<tbody>
<tr>
<td>site</td>
<td>175 176 228 232 274 278</td>
<td>175 176 228 232 274 278</td>
<td>175 176 228 232 274 278</td>
<td>175 176 228 232 274 278</td>
<td>175 176 228 232 274 278</td>
</tr>
</tbody>
</table>

enriched in selection

WT

depleted in selection
**A** Differential selection distribution for HA variants relative to basal 37 °C and relative to basal 39 °C.

**B** Correlation: increased temperature vs. XBP1s induction

\[ r = -0.7675 \]
\[ p < 0.0001 \]

**C** Correlation: increased temperature vs. ATF6f/XBP1s induction

\[ r = -0.7412 \]
\[ p < 0.0001 \]
Correlation: XBP1s induction vs. ATF6f/XBP1s induction

\[ r = 0.7265 \]

\[ p < 0.0001 \]

diffsel (ATF6f/XBP1s 39 °C vs. Basal 39 °C)

diffsel (XBP1s 39 °C vs. Basal 39 °C)
Divergent fitness effects:
Increased temperature and UPR induction

- Increased temperature
- UPR induction

- 0.5
- -0.5
- 1.0
- 1.5
- 2.0

- P228A
- N497G
- P175G
- K176F
- L456D
- K169G
- L456R

- A90D
- K136N
- S153E

- WT HA influenza
- mutant HA influenza

- Basal
- +XBP1s or +XBP1s/ATF6f
- 37 °C
- 39 °C

- quantified diff sel

- quantified diff sel

- DMS |diff sel| - diff sel

- quantified diff sel

- quantified diff sel

- validate
- not validated
- synonymous

- 1:1

- A

- B
A Validated variants: divergent behavior upon XBP1s or ATF6f/XBP1s induction compared to increased temperature

B Not validated variants

C Synonymous controls

Key for A–C:

Induction of UPR:
- +XBP1s 37 °C rel. Basal 37 °C
- +XBP1s 39 °C rel. Basal 39 °C
- +ATF6f/XBP1s 37 °C rel. Basal 37 °C
- +ATF6f/XBP1s 39 °C rel. Basal 39 °C

Increased temperature:
- Basal 39 °C rel. Basal 37 °C
- +XBP1s 39 °C rel. +XBP1s 37 °C
A

<table>
<thead>
<tr>
<th>secondary structure</th>
<th>functional/structural role</th>
<th>conservation</th>
<th>surface accessibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>all sites</td>
<td>helix</td>
<td>ns</td>
<td>*** ** *** ** **</td>
</tr>
<tr>
<td></td>
<td>strand</td>
<td>*** ***</td>
<td>*** *** *** ***</td>
</tr>
<tr>
<td></td>
<td>loop</td>
<td>ns</td>
<td>*** *** *** ***</td>
</tr>
<tr>
<td>receptor binding</td>
<td>ns</td>
<td>**</td>
<td>*** *** *** ***</td>
</tr>
<tr>
<td>sites</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>antigenic sites</td>
<td>ns</td>
<td>**</td>
<td>*** *** *** ***</td>
</tr>
<tr>
<td>glyc. sequon/</td>
<td>ns</td>
<td>**</td>
<td>*** *** *** ***</td>
</tr>
<tr>
<td>disulfide bond</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>conserved</td>
<td>ns</td>
<td>**</td>
<td>*** *** *** ***</td>
</tr>
<tr>
<td>variable</td>
<td>ns</td>
<td>**</td>
<td>*** *** *** ***</td>
</tr>
<tr>
<td>buried</td>
<td>ns</td>
<td>**</td>
<td>*** *** *** ***</td>
</tr>
<tr>
<td>exposed</td>
<td>ns</td>
<td>**</td>
<td>*** *** *** ***</td>
</tr>
</tbody>
</table>

Key for A–C: Basal 39 °C rel. Basal 37 °C +ATF6f/XBP1s 39 °C rel. Basal 39 °C

B

Site conservation versus net diffsel

<table>
<thead>
<tr>
<th>net site diffsel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basal 39 °C rel. Basal 37 °C +ATF6f/XBP1s 39 °C rel. Basal 39 °C</td>
</tr>
</tbody>
</table>

C

Site surface accessibility versus net diffsel

<table>
<thead>
<tr>
<th>net site diffsel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basal 39 °C rel. Basal 37 °C +ATF6f/XBP1s 39 °C rel. Basal 39 °C</td>
</tr>
</tbody>
</table>
A

B

key for A,B:

- net site diff sel (mutational tolerance)

Basal 39 °C
rel. Basal 37 °C

+XBP1s 39 °C
rel. Basal 39 °C

P175, K176
T499

K170
P175, K176
K170
P175, K176
N497
T499
N497
T499

globular head domain
inter-membrane stem domain
B \[K170\] **globular head domain**

**inter-membrane stem domain**

**Basal 39 °C**

rel. Basal 37 °C

**+ATF6f/XBP1s 39 °C**

rel. Basal 39 °C

**P175, K176**

**T499**

**N497**

**key for A,B:**

- **net site diffsel** (mutational tolerance)

---

A **B**