# The endoplasmic reticulum proteostasis network profoundly shapes the protein sequence space accessible to HIV envelope

**Short Title: ER proteostasis shapes HIV envelope sequence space**

Jimin Yoon,[1,2] Emmanuel E. Nekongo,[1,2] Jessica E. Patrick,[2] Tiffani Hui,[3] Angela M. Phillips,[2,4] Anna I. Ponomarenko,[2] Samuel J. Hendel,[2] Rebecca M. Sebastian,[2] Yu Meng Zhang,[2] Vincent L. Butty,[5] C. Brandon Ogbunugafor,[6] Yu-Shan Lin,[3] Matthew D. Shoulders[2,*]

[1]These authors contributed equally to this work.
[2]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[3]Department of Chemistry, Tufts University, Medford, Massachusetts, USA
[4]Current Address: Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, USA
[5]BioMicro Center, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[6]Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut, USA
[*]To whom correspondence may be addressed:
Matthew D. Shoulders
Department of Chemistry
Massachusetts Institute of Technology
77 Massachusetts Avenue, 16-573A
Cambridge, MA 02139
Phone: (617)452-3525
Email: mshoulde@mit.edu

30 **Non-Standard Abbreviations**

31 Env: human immunodeficiency virus-1 envelope protein; DAX (DHFR.ATF6 XBP1s): a cell line that includes

32 the DHFR.ATF6 fusion protein and tetracycline-inducible XBP1s constructs

## Abstract

The sequence space accessible to evolving proteins can be enhanced by cellular chaperones that assist biophysically defective clients in navigating complex folding landscapes. It is also possible, at least in theory, for proteostasis mechanisms that promote strict quality control to greatly constrain accessible protein sequence space. Unfortunately, most efforts to understand how proteostasis mechanisms influence evolution rely on artificial inhibition or genetic knockdown of specific chaperones. The few experiments that perturb quality control pathways also generally modulate the levels of only individual quality control factors. Here, we use chemical genetic strategies to tune proteostasis networks via natural stress response pathways that regulate the levels of entire suites of chaperones and quality control mechanisms. Specifically, we upregulate the unfolded protein response (UPR) to test the hypothesis that the host endoplasmic reticulum (ER) proteostasis network shapes the sequence space accessible to human immunodeficiency virus-1 (HIV) envelope (Env) protein. Elucidating factors that enhance or constrain Env sequence space is critical because Env evolves extremely rapidly, yielding HIV strains with antibody and drug escape mutations. We find that UPR-mediated upregulation of ER proteostasis factors, particularly those controlled by the IRE1-XBP1s UPR arm, globally reduces Env mutational tolerance. Conserved, functionally important Env regions exhibit the largest decreases in mutational tolerance upon XBP1s induction. Our data indicate that this phenomenon likely reflects strict quality control endowed by XBP1s-mediated remodeling of the ER proteostasis environment. Intriguingly, and in contrast, specific regions of Env, including regions targeted by broadly neutralizing antibodies, display enhanced mutational tolerance when XBP1s is induced, hinting at a role for host proteostasis network hijacking in potentiating antibody escape. These observations reveal a key function for proteostasis networks in decreasing instead of expanding the sequence space accessible to client proteins, while also demonstrating that the host ER proteostasis network profoundly shapes the mutational tolerance of Env in ways that could have important consequences for HIV adaptation.

56 **Introduction**

57 Protein mutational tolerance is constrained by the biophysical properties of the evolving protein. Selec-
58 tion to maintain proper protein folding and structure purges a large number of otherwise possible mutations that
59 could be functionally beneficial (1-5). It is no surprise, then, that cellular proteostasis networks play a key role
60 in defining the protein sequence space accessible to client proteins (6-17). Much attention has been given to the
61 phenomenon of chaperones increasing the sequence space accessible to their client proteins, likely by promot-
62 ing the folding of protein variants with biophysically deleterious amino acid substitutions (7-11). Most efforts in
63 this area have focused specifically on how the activities of the heat shock proteins Hsp90 and Hsp70 can expand
64 protein sequence space, in part owing to the availability of specific inhibitors that enable straightforward com-
65 parative studies of protein evolution in the presence versus the absence of folding assistance.

66 In contrast to chaperones increasing sequence space, one might anticipate that protein folding quality
67 control factors would constrain the sequence space accessible to evolving client proteins. For example, promot-
68 ing the rapid degradation and removal of slow-folding or aberrantly folded protein variants could cut off other-
69 wise accessible evolutionary trajectories (16-18), especially if those variants might have still maintained some
70 level of function if instead allowed to persist in the cellular environment. Unfortunately, efforts to understand
71 potential contributions of quality control in shaping protein sequence space are limited. This gap in understand-
72 ing is particularly problematic because natural cellular mechanisms to remodel proteostasis networks function
73 via stress-responsive transcription factors (19, 20), rather than via inhibition or upregulation of individual chap-
74 erones. These transcription factors tune the levels of both chaperones and quality control mechanisms simulta-
75 neously. Such mechanisms may potentially compete in how they impact sequence space of various evolving
76 client proteins.

77 Here, we evaluated whether and how the unfolded protein response (UPR)-regulated endoplasmic retic-
78 ulum (ER) proteostasis network influences the sequence space accessible to membrane proteins processed by
79 the secretory pathway. In particular, we used chemical genetic control of the UPR to broadly modulate the com-
80 position of the ER proteostasis network, and then used deep mutational scanning (DMS) to assess how such

81 perturbations alter accessible client protein sequence space. We chose human immunodeficiency virus-1 (HIV)
82 envelope (Env), a trimeric surface glycoprotein that is folded and quality-controlled by the ER, as our model
83 client protein. We selected Env because its rapid evolution during HIV infections plays a critical role in HIV
84 developing drug and host-cell antibody resistance (21-23). Additionally, Env interacts extensively with various
85 components of the ER proteostasis network, including the ER chaperones calnexin (24) and calreticulin (25),
86 binding immunoglobulin protein (BiP) (26), and ER alpha-mannosidase to initiate ER-associated degradation
87 (ERAD) (27, 28), suggesting the strong potential for the host ER proteostasis network to shape Env's accessible
88 sequence space.

89 Importantly, recent work has revealed that the cellular proteostasis network can indeed impact the se-
90 quence space of not just endogenous client proteins, but also viral proteins that hijack their host's proteostasis
91 machinery (29-33). This relationship has critical evolutionary and therapeutic implications, because mutational
92 tolerance is directly associated with the ability of a virus to evade the host's innate and adaptive immune re-
93 sponses, as well as antiviral drugs (34-40). Early work in this area focused on how viruses like influenza and
94 poliovirus hijack the host's heat shock response-regulated cytosolic chaperones to enhance their mutational tol-
95 erance (29-31). More recently, we discovered that host UPR-mediated upregulation of the ER proteostasis net-
96 work increases the mutational tolerance of influenza A hemagglutinin specifically at febrile temperatures (32).
97 Aside from that hemagglutinin work, no comprehensive studies testing the influence of the ER proteostasis net-
98 work on client protein evolution, whether viral or endogenous, are available.

99 In this study, we used chemical genetic tools to specifically induce the inositol-requiring enzyme-1 / X-
100 box binding protein-1 spliced (IRE1-XBP1s) and the activating transcription factor 6 (ATF6) transcriptional
101 arms of the UPR separately or in tandem (41). This approach provided user-defined modulation of the composi-
102 tion of the host's ER proteostasis network that mimics the cell's natural stress response. We observed that the
103 resulting distinct host environments caused a global decrease in Env mutational tolerance, particularly upon
104 XBP1s-mediated enhancement of the ER proteostasis environment. In addition, we observed that sites with dif-
105 ferent structural or functional roles responded differently to UPR upregulation. For example, conserved regions

106 of Env exhibited an especially strong reduction in mutational tolerance, while a number of sites targeted by
107 broadly neutralizing antibodies displayed an increase in mutational tolerance.

108 This work demonstrates for the first time, to our knowledge, that combined upregulation of chaperones
109 and quality control factors can actually greatly decrease the mutational tolerance of a client protein. It also pro-
110 vides experimental evidence that the host ER proteostasis network profoundly shapes the protein sequence
111 space available to viral membrane proteins and, critically, that the details of the interaction vary from one pro-
112 tein to another – and even within different regions of the same protein.

113 ## Results

114 **Chemical genetic control of ER proteostasis network composition during HIV infection**

115 We began by generating a cell line in which HIV could robustly replicate and we could chemically in-
116 duce the UPR's IRE1-XBP1s and ATF6 transcriptional responses separately or simultaneously, in an ER stress-
117 independent manner. We sought ER stress-independent induction of these transcription factors rather than
118 stress-mediated, global UPR induction, owing to the pleiotropic effects of chemical stressors and the non-physi-
119 ologic, highly deleterious consequences of inducing high levels of protein misfolding in the secretory pathway
120 (19, 32, 41, 42). We selected the IRE1-XBP1s and ATF6 arms of the UPR for chemical control because, in con-
121 trast to the protein kinase R-like ER kinase arm of the UPR that functions largely through translational attenua-
122 tion, they are the key pathways responsible for defining levels of ER chaperones and quality control factors (20,
123 41, 43) likely to influence Env folding, degradation, and secretion.

124 To allow for robust replication of HIV, we chose human T cell lymphoblasts (SupT1 cells) as the host
125 cells. SupT1 cells support high levels of HIV replication in cell culture, likely due to the lack of cytidine deami-
126 nase activity that can cause hypermutation of HIV DNA (44). Moreover, infection with HIVeGFP/VSV-G virus
127 or HIV itself does not alter the expression of UPR-controlled genes in SupT1 cells (45, 46). To attain user con-
128 trol of the IRE1-XBP1s pathway and ATF6 transcriptional response in these cells, we used a previously de-
129 scribed method of stable cell line engineering (41) (detailed in the Methods). Briefly, the XBP1s transcription
130 factor was placed under control of the tetracycline receptor, and induced by treatment with doxycycline (dox).
131 Orthogonally, the active form of the ATF6 transcription factor was fused to an *Escherichia coli* dihydrofolate
132 reductase (DHFR)-based destabilizing domain, and induced by treatment with trimethoprim (TMP). We termed
133 the resulting engineered cells SupT1$^{DAX}$ cells (**Fig 1A**), with the DAX signifier indicating the inclusion of both
134 the DHFR.ATF6 and XBP1s constructs.

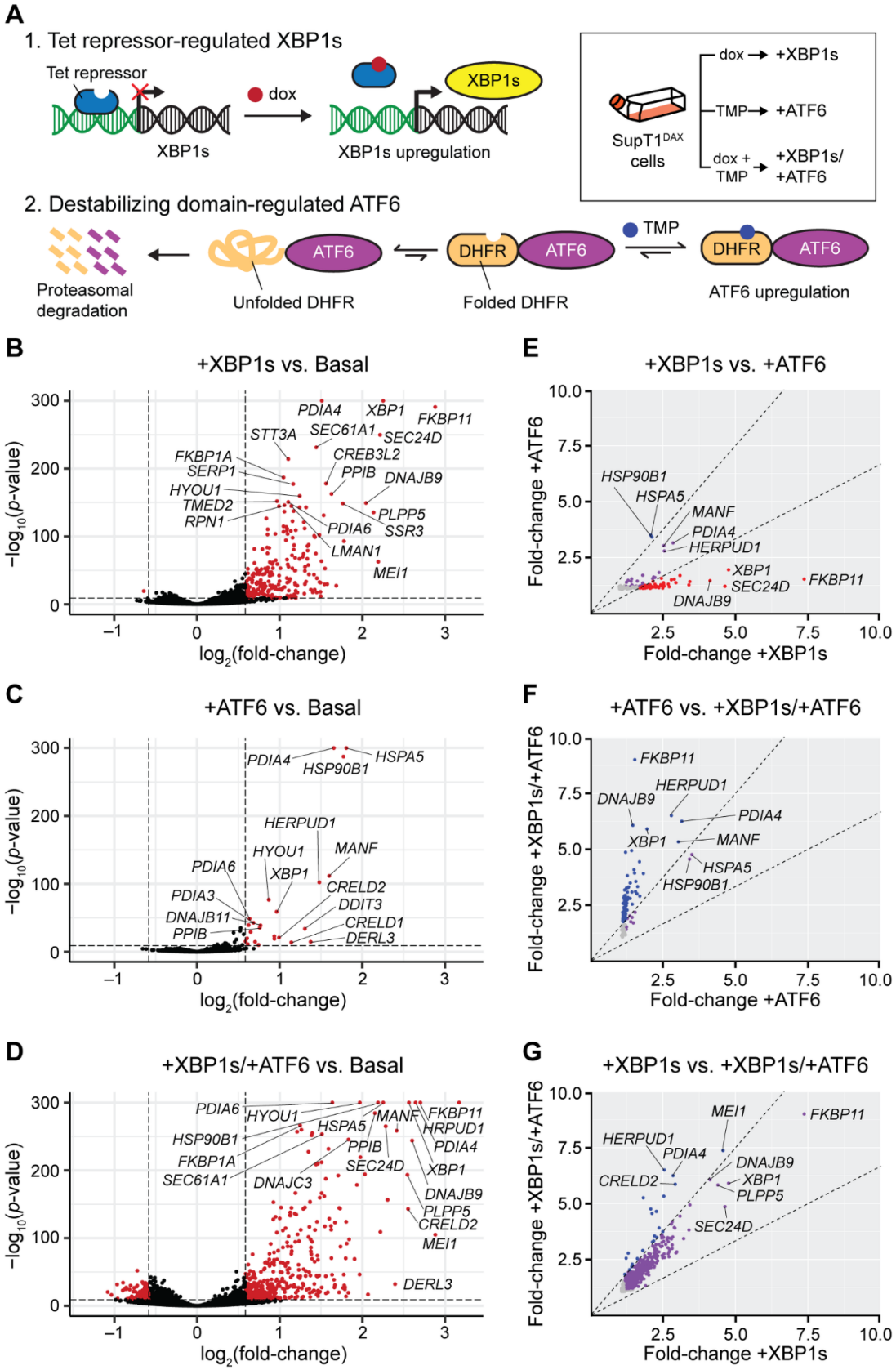



**Fig 1. Stress-independent induction of XBP1s, ATF6, or XBP1s and ATF6 creates four distinct ER proteostasis environments in SupT1$^{DAX}$ cells (basal, +XBP1s, +ATF6, +XBP1s/+ATF6).**

138 (**A**) Chemical genetic strategy to orthogonally regulate XBP1s and ATF6 in SupT1$^{DAX}$ cells. (**B–D**) RNA-Seq analysis of
139 the transcriptomic consequences of (**B**) XBP1s, (**C**) ATF6, and (**D**) XBP1s/ATF6 induction. Transcripts that were differen-
140 tially expressed under each condition based on a >1.5-fold change in expression level (for dox-, TMP-, or dox and TMP-
141 treated versus vehicle-treated cells) and a non-adjusted *p*-value $<10^{-10}$ are separated by dashed lines and plotted in red, with
142 select transcripts labeled. The lowest nonzero *p*-value recorded was $10^{-291}$; therefore, *p*-values = 0 were replaced with *p*-
143 values = $1.00 \times 10^{-300}$ for plotting purposes. Transcripts for which *p*-values could not be calculated owing to extremely low
144 expression or noisy count distributions were excluded from plotting. (**E–G**) Comparison of transcript fold-change upon (**E**)
145 +XBP1s versus +ATF6 (**F**) +ATF6 versus +XBP1s/+ATF6, and (**G**) +XBP1s versus +XBP1s/+ATF6 remodeling of the
146 ER proteostasis network. Only transcripts with false discovery rate-adjusted *p*-value <0.05 and fold-increase >1 in both of
147 the indicated conditions are plotted. Dashed lines indicate a 1.5-fold filter to assign genes as selectively induced by the
148 proteostasis condition on the *x*-axis (red), *y*-axis (blue), or lacking selectivity (purple). Transcripts with fold-increase <1.2
149 in either proteostasis environment are colored in grey to indicate low differential expression. The complete RNA-seq dif-
150 ferential expression analysis is provided in **S1 Data**.
151

152 With stably engineered SupT1$^{DAX}$ cells in hand, we anticipated that we could create four distinct ER
153 proteostasis environments (basal, XBP1s-induced ATF6-induced, and XBP1s/ATF6 co-induced) to assess po-
154 tential consequences for Env mutational tolerance. We induced the XBP1s and ATF6 transcriptional responses
155 in SupT1$^{DAX}$ cells, either separately or together, and evaluated resultant changes in the transcriptome using
156 RNA-Seq (**S1 Data**). We applied gene set enrichment analysis (47) to the RNA-Seq results using the MSigDB
157 c5 collection, and found that gene sets related to ER stress, Golgi trafficking, and ERAD were highly enriched
158 upon induction of XBP1s, ATF6, and co-induction of XBP1s and ATF6 (**S2 Data**). In contrast, gene sets that
159 serve as markers of other stress responses (e.g., the heat shock response) were not enriched, consistent with a
160 highly selective, stress-independent induction of UPR transcriptional responses.

161 Comparing the resulting transcriptomes, we observed significant and substantial upregulation of 223
162 transcripts upon XBP1s induction (+XBP1s), 24 transcripts upon ATF6 induction (+ATF6), and 436 transcripts
163 upon co-induction of XBP1s and ATF6 (+XBP1s/+ATF6) (**Fig 1B–D**). For all three treatment conditions, the
164 upregulated transcripts were strongly biased towards known UPR-regulated components of the ER proteostasis
165 network.

166 To analyze the extent to which these three perturbations (+XBP1s, +ATF6, and +XBP1s/+ATF6) en-
167 gendered unique ER proteostasis environments, we cross-compared the mRNA fold-changes owing to each
168 treatment (**Fig 1E–G**). Transcripts known to be targeted primarily by XBP1s were strongly upregulated upon
169 dox treatment (e.g., *SEC24D* and *DNAJB9*), whereas transcripts known to be targeted primarily by ATF6 were
170 more strongly upregulated upon TMP treatment (e.g., *HSP90B1* and *HSPA5*) (**Fig 1E**) (41, 48, 49). We used
171 immunoblotting to confirm successful induction of these pathways, observing selective protein-level induction
172 of the XBP1s target Sec24D upon dox treatment versus selective induction of the ATF6 target BiP (*HSPA5*)
173 upon TMP treatment (**S1 Fig**). XBP1s induction caused an extensive remodeling of the entire ER proteostasis
174 network, whereas ATF6 induction resulted in targeted upregulation of just a select subset of ER proteostasis
175 factors, consistent with prior work showing that ATF6 induction causes upregulation of fewer transcripts than
176 XBP1s (41, 49). Notably, the combined induction of XBP1s and ATF6 provided access to a third environment
177 where specific transcripts (e.g., genes known to be targets of XBP1s and ATF6 heterodimers, such as *HER-*
178 *PUD1*) were more strongly upregulated than upon the single induction of either transcription factor (**Fig 1F** and
179 **1G**) (41, 50, 51). Taken together, our RNA-Seq results show that we can access four distinctive ER proteostasis
180 environments for Env mutational tolerance experiments via chemical genetical control of XBP1s and ATF6 (ba-
181 sal, +XBP1s, +ATF6, and +XBP1s/+ATF6).
182 We assessed whether these perturbations of the ER proteostasis environment had deleterious effects on
183 cell viability or restricted HIV replication, as we had previously observed inhibition of HIV replication upon
184 upregulation of the heat shock response (52). To address the former, we induced XBP1s and ATF6, individually
185 or simultaneously, in SupT1$^{DAX}$ cells and measured resazurin metabolism 72 h post-drug treatment (**S2A Fig**).
186 We observed that induction of XBP1s and ATF6, either separately or simultaneously, did not alter the metabolic
187 activity of SupT1$^{DAX}$ cells, consistent with no deleterious effects on cell viability. To address the latter, we used
188 the TZM-bl assay to quantify HIV infectious titer (**S2B Fig**). Specifically, we used TZM-bl reporter cells con-
189 taining the *E. coli* β-galactosidase gene under the control of an HIV long-terminal repeat sequence (53). When
190 these cells are infected with HIV, the HIV Tat transactivation protein induces expression of β-galactosidase,

191 which cleaves the chromogenic substrate (X-Gal) and causes infected cells to appear blue in color. The infec-
192 tious titer increased marginally by approximately 3.5-fold when XBP1s was induced, either alone or together
193 with ATF6. Induction of ATF6 alone did not affect HIV infectious titer. Thus, ER proteostasis network pertur-
194 bation via XBP1s and/or ATF6 induction did not deleteriously impact HIV replication.
195

196 **Env deep mutational scanning in four distinct host ER proteostasis environments**

197 We next applied DMS to Env to test our hypothesis that the composition of the host's ER proteostasis
198 network plays a central role in determining the mutational tolerance of Env. For this purpose, we employed a
199 previously developed set of three replicate Env proviral plasmid libraries (22), created by introducing random
200 codon mutations at amino acid residues 31–702 of the Env protein (note that the HXB2 numbering scheme (54)
201 is used throughout). Briefly, the library was generated using a previously described technique that uses pools of
202 primers containing a random NNN nucleotide sequence at the codon of interest, and mutations are introduced
203 via iterative rounds of low-cycle PCR (55). This technique generates multi-nucleotide (e.g., gca → gAT) as well
204 as single nucleotide (e.g., gca → gAa) codon mutations, thereby introducing mutations at the codon level rather
205 than at the nucleotide level (22, 55). The N-terminal signal peptide and the C-terminal cytoplasmic tail of Env
206 were excluded from mutagenesis owing to their dramatic impact on Env expression and/or HIV infectivity (22).
207 We generated biological triplicate viral libraries from these mutant Env plasmid libraries by transfecting
208 the plasmid libraries into HEK293T cells and then harvesting the passage 0 (p0) viral supernatant after 4 d.
209 Deep sequencing of the three p0 viral libraries showed that 74% of all possible amino-acid substitutions were
210 observed at least three times in each of the triplicate libraries, and 98% of all possible substitutions were ob-
211 served at least three times in at least one of the triplicate libraries, consistent with prior work (22, 36). Mutations
212 that were not included in the viral libraries were dispersed throughout the sequence and did not correspond to
213 specific regions of structural or functional importance (**S3 Fig**). To establish a genotype–phenotype link, we
214 passaged the p0 transfection supernatants in SupT1 cells at a very low multiplicity of infection (MOI) of 0.005
215 infectious virions/cell. We next performed batch competitions of each individual Env viral library in SupT1$^{DAX}$

216 cells in each of the four different ER proteostasis environments: basal, +XBP1s, +ATF6, and +XBP1s/+ATF6

217 (**Fig 2A**). Briefly, SupT1$^{DAX}$ cells were treated with vehicle, dox, TMP, or both dox and TMP to generate the

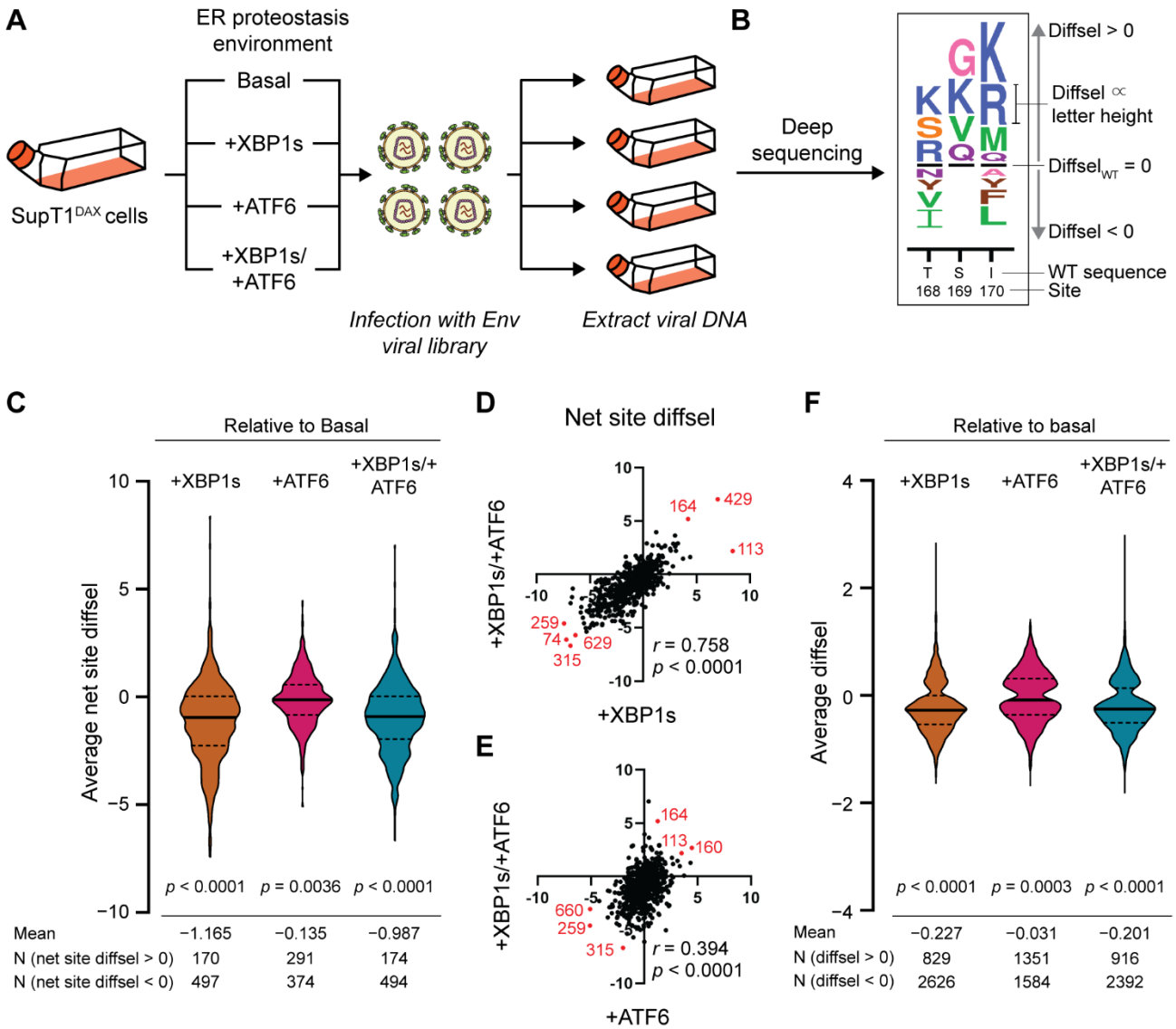
218 intended ER proteostasis environment, followed by infection with p1 viral supernatant at a MOI of 0.005 infec-

219 tious virions/cell. We used this MOI to minimize co-infection of individual cells and thereby maintain the geno-

220 type–phenotype link. Non-integrated viral DNA was extracted and *Env* amplicons were generated by PCR (22).

221 Finally, we deep-sequenced the amplicons using barcoded-subamplicon sequencing (**S4 Fig**) and analyzed the

222 sequencing reads using the dms_tools2 suite (https://jbloomlab.github.io/dms_tools2/) (56, 57).




223

224 **Fig 2. Upregulation of the host cell's ER proteostasis environment generally reduces mutational tolerance across the**

225 **Env protein sequence.**

226 **(A)** Scheme for deep mutational scanning of Env in four distinct ER proteostasis environments (basal, +XBP1s, +ATF6,
227 +XBP1s/+ATF6). SupT1$^{DAX}$ cells were pre-treated with DMSO (Basal), dox (+XBP1s), TMP (+ATF6), or both dox and
228 TMP (+XBP1s/+ATF6) 18 h prior to infection with biological triplicate Env viral libraries. 4 d post-infection, cells were
229 harvested, and non-integrated viral DNA was sequenced to quantify diffsel of Env variants. **(B)** Diffsel for each amino-acid
230 variant can be visualized in a sequence logo plot. The black horizontal lines represent the diffsel for the wild-type amino
231 acid at that site, and the height of the amino-acid letter abbreviations is proportional to the diffsel of that variant in the
232 remodeled ER proteostasis environment relative to the basal environment. Variants that are relatively enriched in the indi-
233 cated ER proteostasis environment (positive diffsel) are located above the black horizontal line. Variants that are relatively
234 depleted in the indicated ER proteostasis environment (negative diffsel) are located below the black horizontal line. (**C**) Net
235 site diffsel for all Env sites in three perturbed ER proteostasis environments, averaged over biological triplicates. The black
236 horizontal lines on the violin plots indicate the median (solid line) or the first and the third quartiles (dashed lines) of the
237 distribution. The significance of deviation from null (net site diffsel = 0, no selection) was tested using a one-sample *t*-test,
238 with two-tailed *p*-values shown. The mean of the distribution, as well as the number of sites with net site diffsel >0 or <0,
239 are listed below the distribution. (**D–E**) Correlation for net site diffsel values between (**D**) +XBP1s/+ATF6 versus +XBP1s
240 and (**E**) +XBP1s/+ATF6 versus +ATF6, normalized to the basal proteostasis environment. Pearson correlation coefficients
241 *r* and corresponding *p*-values are shown. Select sites with highly positive or highly negative net site diffsel values in both
242 proteostasis environments are marked in red and labeled with site numbers. (**F**) Diffsel for individual Env variants in three
243 perturbed ER proteostasis environments, averaged over biological triplicates. The black horizontal lines on the violin plots
244 indicate the median (solid line) or the first and the third quartiles (dashed lines) of the distribution. The significance of
245 deviation from null (diffsel = 0, no selection) was tested using a one-sample *t*-test, with two-tailed *p*-values shown. The
246 mean of the distribution, as well as the number of sites with diffsel >0 and <0, are listed below the distribution. (**C–F**)
247 Diffsel values are provided in https://github.com/yoon-jimin/2021_HIV_Env_DMS.
248

249 To identify amino-acid variants that were differentially enriched or depleted in a given ER proteostasis
250 selection condition (+XBP1s, +ATF6, or +XBP1s/+ATF6) relative to the basal ER proteostasis environment,
251 we quantified differential selection (diffsel) (**Fig 2B**). Diffsel was calculated by taking the logarithm of the vari-
252 ant's enrichment in the selection condition relative to its enrichment in the basal ER proteostasis network condi-
253 tion (57). For example, if a variant exhibited positive diffsel in +XBP1s (selection) versus basal (mock), it
254 would indicate that the variant was more enriched relative to the wild-type amino acid in the +XBP1s condition
255 compared to the basal condition. In addition, to decipher reliable signal from experimental noise, we filtered the
256 DMS data using a previously described and validated two-step strategy (32). First, we removed variants that
257 were not present in all three pre-selection replicate viral libraries. That is, we eliminated even those variants that

258 were strongly enriched or depleted in two replicates if they were not present in the starting library of the third
259 replicate. Second, we removed variants that exhibited diffsel in opposite directions in any of the biological trip-
260 licates. Using the second filter, we typically removed variants that were minimally affected by the selection,
261 displaying slightly positive diffsel values in one replicate but slightly negative diffsel values in another. By ap-
262 plying these two filters, we were able to focus subsequent analyses only on Env variants that exhibited robust,
263 reproducible diffsel across biological triplicates of the same ER proteostasis network conditions (out of 12,787
264 theoretically possible non-wild-type variants; 3,455 variants for +XBP1s (27%), 2,935 variants for +ATF6
265 (23%), and 3,308 variants for +XBP1s/+ATF6 (26%)).
266

267 **XBP1s induction causes a strong net decrease in the mutational tolerance of Env, consistent with en-**
268 **hanced quality control of biophysically defective variants**

269 To evaluate our hypothesis that the composition of the host's ER proteostasis network critically shapes
270 Env mutational tolerance, we first analyzed the 'net site diffsel' in each host ER proteostasis environment. Net
271 site diffsel is the sum of individual mutational diffsel values for a given Env site. Thus, a positive net site diffsel
272 indicates that mutational tolerance at a given Env site is quantitatively increased in an enhanced host ER proteo-
273 stasis environment relative to the basal ER proteostasis environment. In contrast, a negative net site diffsel indi-
274 cates that mutational tolerance is decreased in an enhanced host ER proteostasis environment. For example, the
275 net site diffsel for site 169 (**Fig 2B**) would be the sum of the diffsel values for G, K, V, and Q, which would be
276 positive, and therefore, we would conclude that the overall mutational tolerance, as defined here, increased at
277 site 169.

278 Using the filtered Env DMS datasets, we calculated net site diffsel at each Env position averaged across
279 the three biological replicates of our experiment (**Fig 2C**). Strikingly, the +XBP1s ER proteostasis environment
280 globally, substantially, and significantly reduced mutational tolerance across the entire Env protein (mean net
281 site diffsel = −1.165, $p$-value < 0.0001). Co-induction of XBP1s and ATF6 had a similar effect, again substan-
282 tially and significantly reducing Env mutational tolerance (mean net site diffsel = −0.987, $p$-value < 0.0001).

283 The magnitude of absolute mean net site diffsel was approximately 14-fold larger upon XBP1s induction than
284 we previously observed for increased mutational tolerance in influenza hemagglutinin in an XBP1s-activated
285 ER proteostasis environment at 37 °C (32). Thus, Env mutational tolerance is exceptionally sensitive to XBP1s-
286 mediated ER proteostasis network upregulation, to a much greater extent than hemagglutinin. In contrast, the
287 +ATF6 ER proteostasis environment, while still mildly reducing mutational tolerance across Env, had a less
288 substantial global effect (mean net site diffsel = −0.135, *p*-value = 0.0036). The latter result suggests that the
289 reduced Env mutational tolerance observed in the +XBP1s/+ATF6 ER proteostasis environment was largely
290 driven by ER proteostasis factors targeted by XBP1s. Indeed, the Pearson correlation coefficient *r* was substan-
291 tially higher between the net site diffsel values observed in the +XBP1s versus +XBP1s/+ATF6 environments (*r*
292 = 0.758; **Fig 2D**) than between those observed in the +ATF6 versus +XBP1s/+ATF6 (*r* = 0.394; **Fig 2E**). This
293 observation aligns well with our RNA-Seq data, in which we observed substantially more overlap between the
294 ER proteostasis network transcriptome remodeling caused by XBP1s induction versus the co-induction of
295 XBP1s and ATF6, compared to ATF6 induction versus the co-induction of XBP1s and ATF6 (**Fig 1F** and **1G**).
296 It is important to note that, in a net site diffsel analysis, we quantify the relative enrichment of all amino-
297 acid variants combined to assess mutational tolerance at a given Env site. Consequently, a decrease in muta-
298 tional tolerance as measured by net site diffsel could be caused by a single amino-acid variant that was strongly
299 disfavored or, alternatively, by many variants being disfavored relative to wild-type. To test if individual amino-
300 acid variants also reveal a global trend towards reduced mutational fitness, we plotted the individual diffsel val-
301 ues for all Env variants. We again observed reduced mutational fitness of the majority of Env variants whenever
302 XBP1s was induced, indicating that the effect is largely driven by a general loss of mutational tolerance rather
303 than by just a few specific amino-acid variants that were strongly disfavored (**Fig 2F**).
304 The unanticipated and striking decrease in mutational tolerance of Env upon XBP1s induction could po-
305 tentially arise from the fact that XBP1s upregulates both chaperones that assist client protein folding and quality
306 control factors that identify and dispose of defective proteins. We used the Rosetta $\Delta\Delta G$ protocol to predict the
307 energetic consequences of all amino-acid substitutions that were present in our filtered DMS dataset (**S3 Data**)

308 (58). Although there are limitations associated with using the Rosetta cartesian_ddg protocol to predict the ex-
309 act, absolute changes in protein folding free energy upon substitution, the protocol and associated scaling fac-
310 tors can provide relative stability of substitutions and a general classification between destabilizing and stabiliz-
311 ing substitutions (58). Disulfide-bonding cysteine residues, which the Rosetta protocol defines as a feature and
312 disallows substitutions, were excluded from $\Delta\Delta G$ prediction, although substitutions in these disulfide-bonding
313 cysteines can be presumed to be highly destabilizing owing to the critical structural roles of disulfide bonds. To
314 test whether the variants that exhibit negative diffsel values upon XBP1s induction are more destabilizing than
315 those with positive diffsel values, we compared the distribution of predicted $\Delta\Delta G$ for all variants with positive
316 diffsel versus negative diffsel (**Fig 3A** and **3B**). We observed that the variants with negative diffsel on average
317 had moderately higher (more destabilizing) predicted $\Delta\Delta G$ than the variants with positive diffsel (two-sample *t*-
318 test, two-tailed $p$-value $< 0.0001$). To further test if substitutions at mutationally intolerant sites upon XBP1s
319 induction are generally destabilizing, we focused on the 20 most negative and the 20 most positive net site
320 diffsel positions (**Fig 3C** and **3D**). We again found that, overall, substitutions at sites with strongly negative net
321 site diffsel (sites with low mutational tolerance) were much more destabilizing than substitutions at sites with
322 strongly positive net site diffsel (sites with high mutational tolerance).

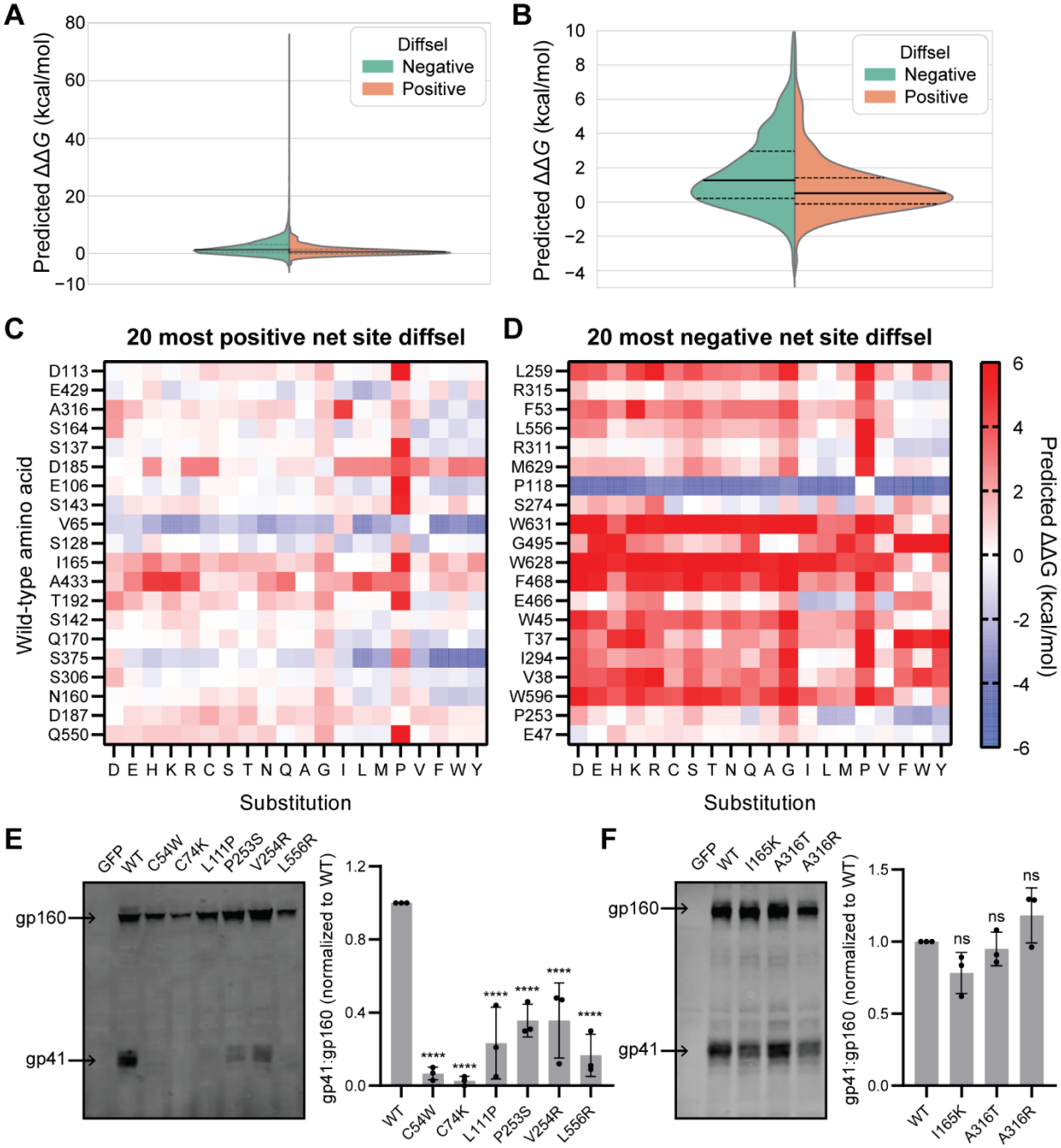



**Fig 3. Env variants displaying negative diffsel upon XBP1s induction tend to be more destabilizing and exhibit greater processing defects than those displaying positive diffsel.** (**A**) Split violin plot depicting the distribution of $\Delta\Delta G$ values predicted using the Rosetta $\Delta\Delta G$ protocol, for all amino-acid substitutions that were present in the filtered DMS dataset for +XBP1s vs. Basal (2379 negative diffsel variants; 756 positive diffsel variants). Dashed lines inside the violins indicate the first and third quartiles, and the solid line inside the violins indicates the median. (**B**) Zoom-in of the violin plot in (**A**) focusing on $\Delta\Delta G < 10$ kcal/mol. (**C, D**) Heatmaps showing the predicted $\Delta\Delta G$ values for all possible amino-acid substitutions at the 20 sites with the most positive net site diffsel (**C**) and the 20 sites with the most negative net site diffsel (**D**), upon XBP1s induction. Substitutions (x-coordinate) are arranged by side-chain properties: negatively charged (D, E), positively charged (H, K R), polar uncharged (C, S, T, N, Q), small non-polar (A, G), aliphatic (I, L, M, P, V), and aromatic

333 (F, W, Y). Wild-type amino acids (y-coordinate) are arranged by rank order of net site diffsel, with (**C**) D113 most positive
334 and (**D**) L259 most negative. Complete $\Delta\Delta G$ values are provided in **S5 Data**. (**E**) Representative immunoblot showing
335 gp160 and gp41 bands for selected variants with negative diffsel upon XBP1s induction (left) and densitometric analysis of
336 gp41:gp160 ratio across biological triplicates (right). (**F**) Representative immunoblot showing gp160 and gp41 bands for
337 selected variants with positive diffsel upon XBP1s induction (left) and densitometric analysis of the gp41:gp160 ratio across
338 biological triplicates (right). For (**E**) and (**F**) statistical significance was calculated by one-way ANOVA, comparing the
339 mean of each variant to the mean of wild type (WT). *, **, ***, and **** represent ANOVA *p*-values of <0.05, <0,01,
340 <0.001, and <0.0001, respectively. Immunoblots of biological triplicates are provided in **S5 Fig** and replicate data values
341 for densitometric analysis are provided in **S6 Data**.
342
343 Rosetta $\Delta\Delta G$ only makes predictions regarding thermodynamic stability, whereas variants can also in-
344 duce defects in the kinetics of folding or proper processing of Env. We next used an experimental approach to
345 assess whether variants displaying negative diffsel values upon XBP1s induction displayed more serious traf-
346 ficking defects than those with positive diffsel values. Since Env is synthesized as a precursor protein (gp160)
347 in the ER and proteolytically cleaved into gp120 and gp41 in the Golgi, Env variants that fail to pass ER quality
348 control would be predicted to result in a lower gp41:gp160 or gp120:gp160 ratio compared to wild-type Env
349 (59-61). We chose six variants with strongly negative diffsel (C54W, C74K, L111P, P253S, V254R, L556R)
350 and three variants with strongly positive diffsel (I165K, A316T, A316R), transfected them into HEK293T cells,
351 and calculated the steady-state ratio of gp41 to gp160. We observed that all the tested Env variants with nega-
352 tive diffsel upon XBP1s induction exhibited lower gp41:gp160 than wild-type Env (**Fig 3E**), while the ratio was
353 only slightly lower or sometimes higher than wild-type Env for variants with positive diffsel (**Fig 3F**). Of note,
354 gp41 bands were nearly undetectable when substitutions were made at disulfide-bonding cysteines, confirming
355 that substitutions at these cysteines do severely disrupt Env trafficking. Together, our Rosetta $\Delta\Delta G$ predictions
356 and experimental data strongly support the hypothesis that the Env variants rendered less fit upon XBP1s induc-
357 tion were more energetically destabilizing and disrupted Env maturation more strongly than the variants that
358 were enriched upon XBP1s induction.
359 While it is known that infection with HIVeGFP/VSV-G virus or HIV itself does not result in UPR up-
360 regulation in SupT1 cells (45, 46), it is possible that that the destabilized or poorly folding variants in our

361 library may significantly misfold in the ER and result in more pronounced UPR activation. To address this pos-
362 sibility, we transfected wild-type Env and Env variants that were strongly negatively selected (C54W, L111P,
363 L556R) into HEK293T cells (instead of SupT1 cells where high-efficiency transfection is not possible) and
364 measured UPR upregulation using real-time PCR (**S6 Fig**). Overall, both the wild-type Env and the three vari-
365 ants displayed UPR signaling equivalent to GFP-transfected cells (negative control), and to a much lower level
366 than the GFP-transfected cells treated with ER stress inducer thapsigargin (Tg; positive control). This result in-
367 dicates that it is unlikely that the destabilized variants in our library activated the UPR above the basal level.

368 In sum, there is a striking decrease in mutational tolerance across much of Env upon XBP1s-mediated
369 remodeling of the host's ER proteostasis network. Although unexpected, this observation is actually quite con-
370 sistent with XBP1s-upregulated quality control factors restricting the available protein sequence space by enact-
371 ing stringent quality control on biophysically defective protein variants. This broad and substantive trend should
372 not, however, mask the fact that numerous sites displayed strongly enhanced mutational tolerance upon not just
373 XBP1s induction but also ATF6 induction (e.g., S164, D113) (**Fig 2C**; N (net site diffsel > 0), **2D**, and **2E**). Fi-
374 nally, it should be noted that although ATF6 induction had minimal global consequences for Env mutational
375 tolerance, there were still a number of sites where reduced net site diffsel (e.g., L259, R315) was observed
376 across all three enhanced ER proteostasis environments (**Fig 2D** and **2E**).

377

378 **Investigation of Env sites and variants most strongly impacted by the host's ER proteostasis network**

379 To visualize the relative fitness of individual amino-acid variants in each host ER proteostasis environ-
380 ment, we generated sequence logo plots across the entire Env sequence (**Fig 4A**, **S7 Fig**, and **S8 Fig**). The rela-
381 tive enrichment for each amino-acid variant (diffsel) was calculated from our filtered data sets by averaging
382 across three biological replicates. The unfiltered, unaveraged full sequence logo plots for each replicate and
383 condition are also provided in https://github.com/yoon-jimin/2021_HIV_Env_DMS.

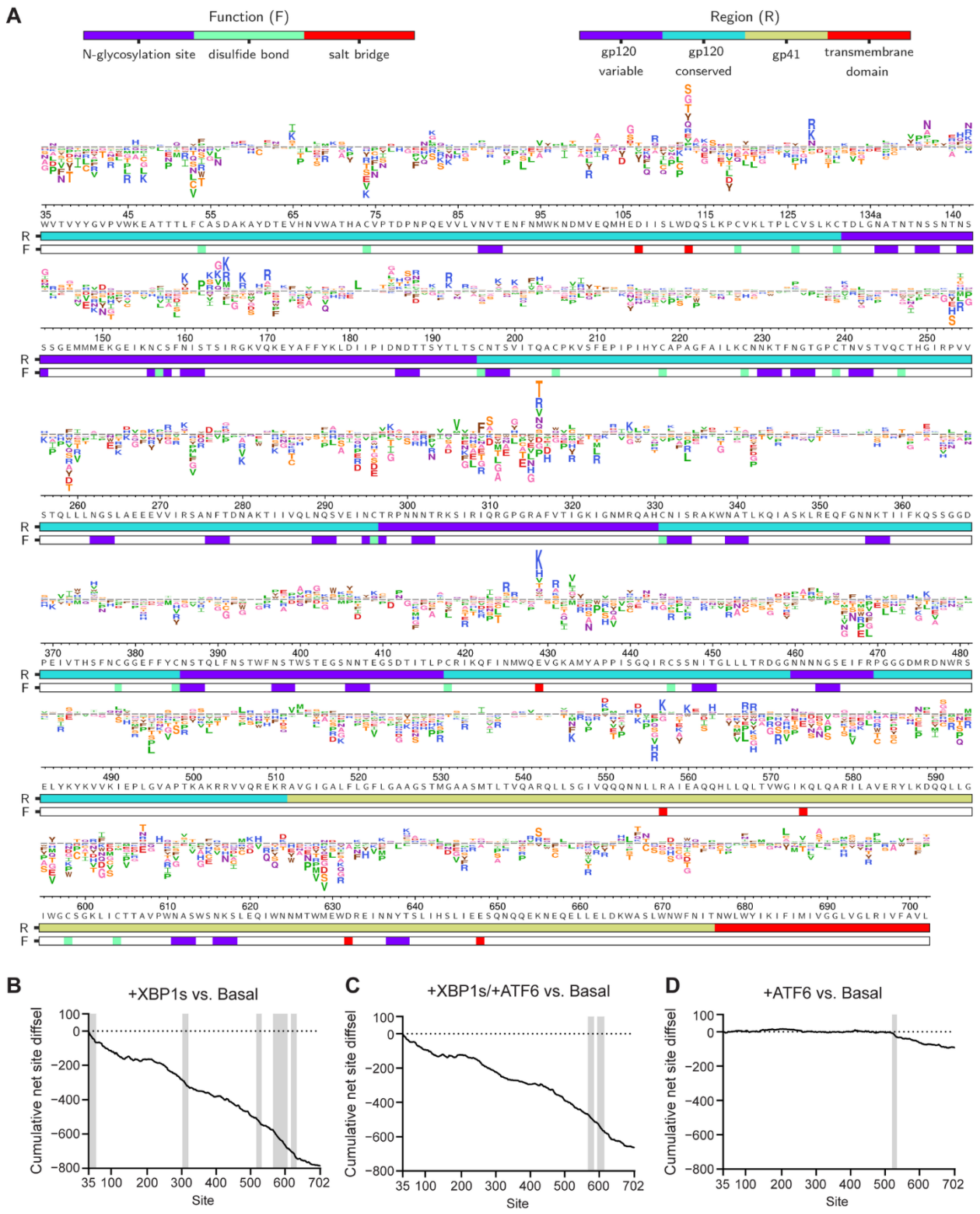



**Fig 4. Diffsel across Env upon remodeling of the host's ER proteostasis network.**

386 (**A**) Logo plot displaying averaged diffsel for +XBP1s normalized to the basal proteostasis environment. The height of the
387 amino-acid abbreviation is proportional to the magnitude of diffsel. The amino-acid abbreviations are colored based on the
388 side-chain properties: negatively charged (D, E; red), positively charged (H, K R; blue), polar uncharged (C, S, T; orange /
389 N, Q; purple), small nonpolar (A, G; pink), aliphatic (I, L, M, P, V; green), and aromatic (F, W, Y; brown). The numbers
390 and letters below the logos indicate the Env site in HXB2 numbering and the identity of the wild-type amino acid for that
391 site, respectively. The color bar below the logos indicates the function (F) that the site is involved in (*N*-glycosylation site
392 (purple), disulfide bond (green), or salt bridge (red)) or the region (R) of Env that the site belongs to (gp120-variable (purple),
393 gp120-conserved (cyan), gp41 (yellow), or transmembrane domain (red); the sites that belong to the five variable loops of
394 gp120 were categorized as 'gp120-variable', and the sites that are not included in the five variable loops were categorized
395 as 'gp120-conserved'). Only variants that were present in all three pre-selection viral libraries and exhibited diffsel in the
396 same direction across all three biological triplicates are plotted here. Diffsel values and unfiltered logo plots for each indi-
397 vidual replicate are provided in https://github.com/yoon-jimin/2021_HIV_Env_DMS. (**B–D**) Cumulative net site diffsel
398 across Env sites for (**B**) +XBP1s, (**C**) +XBP1s/+ATF6, and (**D**) +ATF6, normalized to the basal proteostasis environment.
399 Regions where the decrease in mutational tolerance is particularly prominent are shaded in grey (40–57, 302–319, 517–532,
400 565–607, and 617–633 for (**B**), 567–585 and 594–614 for and (**C**), 520–534 for (**D**)). Cumulative net site diffsel data values
401 are provided in **S8 Data**.
402

403 Several features of these logo plots are immediately noteworthy. First, the global and relatively similar
404 reduction in mutational tolerance caused by XBP1s induction (**Fig 4A**) and co-induction of XBP1s and ATF6
405 (**S7 Fig**) is readily observed. To visualize this phenomenon and highlight specific regions in which the effect
406 size is particularly large, we plotted cumulative net site diffsel against Env sites (**Fig 4B–D**). We observed that
407 the decrease in mutational tolerance was most prominent around the following sites: 40–57, 302–319, 517–532,
408 565–607, and 617–633 when XBP1s was induced alone (**Fig 4A** and **4B**); 567–585 and 594–614 when XBP1s
409 and ATF6 were co-induced (**S7 Fig** and **Fig 4C**), and 520–534 when ATF6 was induced alone (**S8 Fig** and **Fig**
410 **4D**), as indicated by the steeper slopes in those regions. In all three proteostasis environments, sites with strong
411 decreases in mutational tolerance included regions in gp41 (residues 512–702). Second, although the general
412 trend towards reduced mutational tolerance was quite striking, it is also apparent that there are specific positions
413 where either XBP1s- or ATF6-mediated ER proteostasis network enhancement strongly enhanced mutational
414 tolerance at a given site (e.g., D113) or enhanced the fitness of a specific variant (e.g., I309F). We assessed
415 whether this differential impact of ER proteostasis mechanisms was related to surface accessibility of sites, but

did not observe a strong linear correlation between net site diffsel and surface accessibility across Env sites for either the Env monomer or the trimer (**S9 Fig**). Still, we observed that when XBP1s was induced, either alone or together with ATF6, sites that have high surface accessibility were more likely to have positive net site diffsel than sites that have low surface accessibility. Third, the stronger impacts of XBP1s induction compared to ATF6 induction are apparent (**Fig 4A** versus **S8 Fig** and **Fig 4B** versus **4D**).

To assess whether or not the global decrease in mutational tolerance could be attributed to specific structural or functional regions, we calculated the average net site diffsel for individual functional/structural groups. These groups included (1) the transmembrane and soluble domains, (2) the entire gp120 and gp41 subunits, (3) the conserved and variable regions of gp120, where the conserved region is defined as the region that does not belong to the five variable loops of gp120, (4) the five variable loops of gp120 individually (denoted V1–5), (5) regions responsible for viral membrane fusion, and (6) other sites with important functional and structural roles (**Fig 5**, **S10 Fig**, and **S11 Fig**; see corresponding references for assignment of these regions in **S2 Table**).

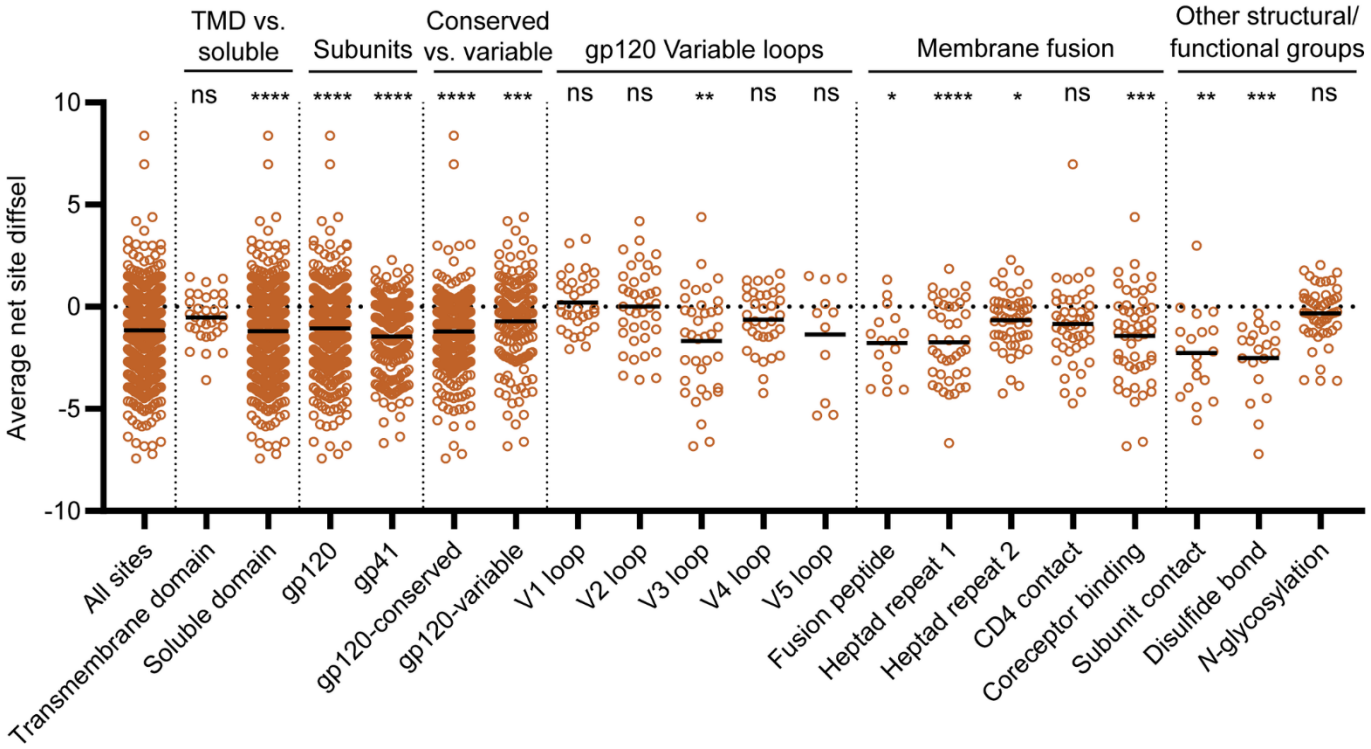



**Fig 5. Impact of XBP1s induction on mutational tolerance varies across Env structural elements.** Average net site diffsel for the +XBP1s ER proteostasis environment normalized to the basal ER proteostasis environment, where the means of the distributions are indicated by black horizontal lines. Sites are sorted by TMD vs. soluble, subunits,

conserved vs. variable regions of gp120, five variable loops of gp120, regions important for membrane fusion, and other structural/functional groups. For 'TMD vs. soluble', all sites that do not belong to the TMD were categorized as 'soluble'. For 'Conserved vs. variable', the sites that belong to the five variable loops of gp120 were categorized as 'gp120-variable', and the sites that are not included in the five variable loops were categorized as 'gp120-conserved'. Significance of deviation from null (net site diffsel = 0, no selection) was tested using a one-sample *t*-test. The derived *p*-values were Bonferroni-corrected for 20 tests and *, **, ***, and **** represent adjusted two-tailed *p*-values of <0.05, <0,01, <0.001, and <0.0001, respectively. Diffsel values are provided in https://github.com/yoon-jimin/2021_HIV_Env_DMS. Assignment of structural regions is provided in **S2 Table**.

We focused first on the consequences of XBP1s induction because the effects were larger than for ATF6 induction and similar to the consequences of dual induction. We examined the mutational tolerance of the trans-membrane domain (TMD) of Env, since recent studies have suggested that the TMDs of other membrane proteins exhibit particularly restricted mutational tolerance (**Fig 5**; 'TMD vs. soluble') (18, 62). We observed a reduction in mutational tolerance for TMD, but the difference was not statistically significant, and the mean net site diffsel for the TMD was less negative than that of the soluble domains. While it is certainly possible that the TMD of Env has highly restricted mutational tolerance, that mutational tolerance (or intolerance) was not particularly altered by XBP1s induction.

We observed a decrease in mutational tolerance for both gp120 and gp41, indicating that XBP1s upregulation impacts both subunits of Env, albeit gp41 more strongly (**Fig 5**; 'Subunits'). Within the gp120 subunit, there was a stronger decrease in mutational tolerance for the regions that did not belong to any variable loops (gp120-conserved) than there was for the variable loops (gp120-variable), although both conserved and variable regions exhibited a loss of mutational tolerance (**Fig 5**; 'Conserved vs. variable'). Among the five variable loops of gp120, the more conserved V3 loop exhibited the strongest negative net site diffsel (**Fig 5**; 'gp120 Variable loops') (63). Further notable within the V3 loop, we observed a particularly large decrease in mutational tolerance for sites that are highly conserved, such as the GPGR motif or the hydrophobic patch whose disruption causes gp120 shedding (**Fig 6A** and **6B)** (64).

To test whether sequence variability correlated with mutational tolerance across the entire Env protein, we plotted net site diffsel against Shannon entropy, which is a measure of sequence variability within Env

sequences of various HIV strains. Indeed, although the linear correlation was not high, 53.6% of positions with high Shannon entropy exhibited increases in mutational tolerance, while only 20.5% of positions did so in conserved positions (**S12 Fig**). These observations suggest that conserved regions in Env generally experience stronger selection pressure when the ER proteostasis network is upregulated than do variable regions.

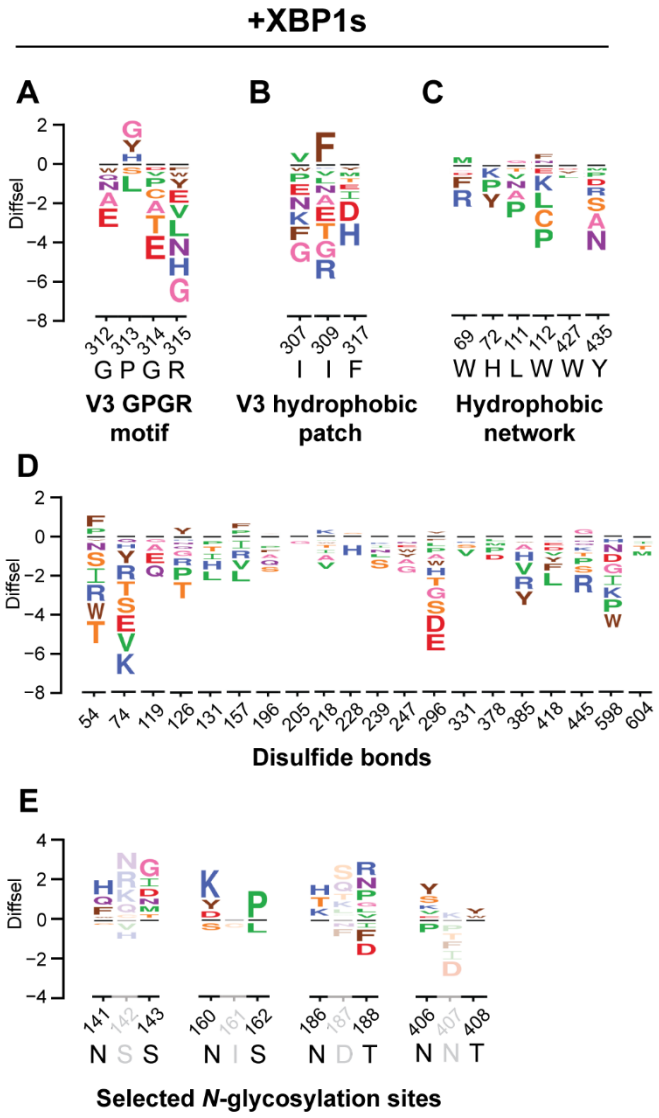



**Fig 6. Diverse functional elements of Env respond differently to XBP1s induction.**
Selected sequence logo plots for the +XBP1s ER proteostasis environment normalized to the basal ER proteostasis environment for (**A**) the conserved GPGR motif of the V3 loop, (**B**) the hydrophobic patch of the V3 loop, (**C**) the hydrophobic network of gp120 important for CD4 binding, (**D**) cysteine residues participating in disulfide bonds, and (**E**) selected *N*-glycosylation sequons (N-X-S/T) that exhibited positive net site diffsel in all three remodeled proteostasis environments. The height of the amino-acid abbreviation corresponds to the magnitude of diffsel. The numbers and letters below the logos indicate the Env site in HXB2 numbering and the wild-type amino acid for that site, respectively. Only variants that were

472 present in all three pre-selection viral libraries and exhibited diffsel in the same direction across the biological triplicates
473 are plotted. All logo plots were generated on the same scale. Diffsel values are provided in https://github.com/yoon-ji-
474 min/2021_HIV_Env_DMS. Assignments of functional regions are provided in **S2 Table**.
475

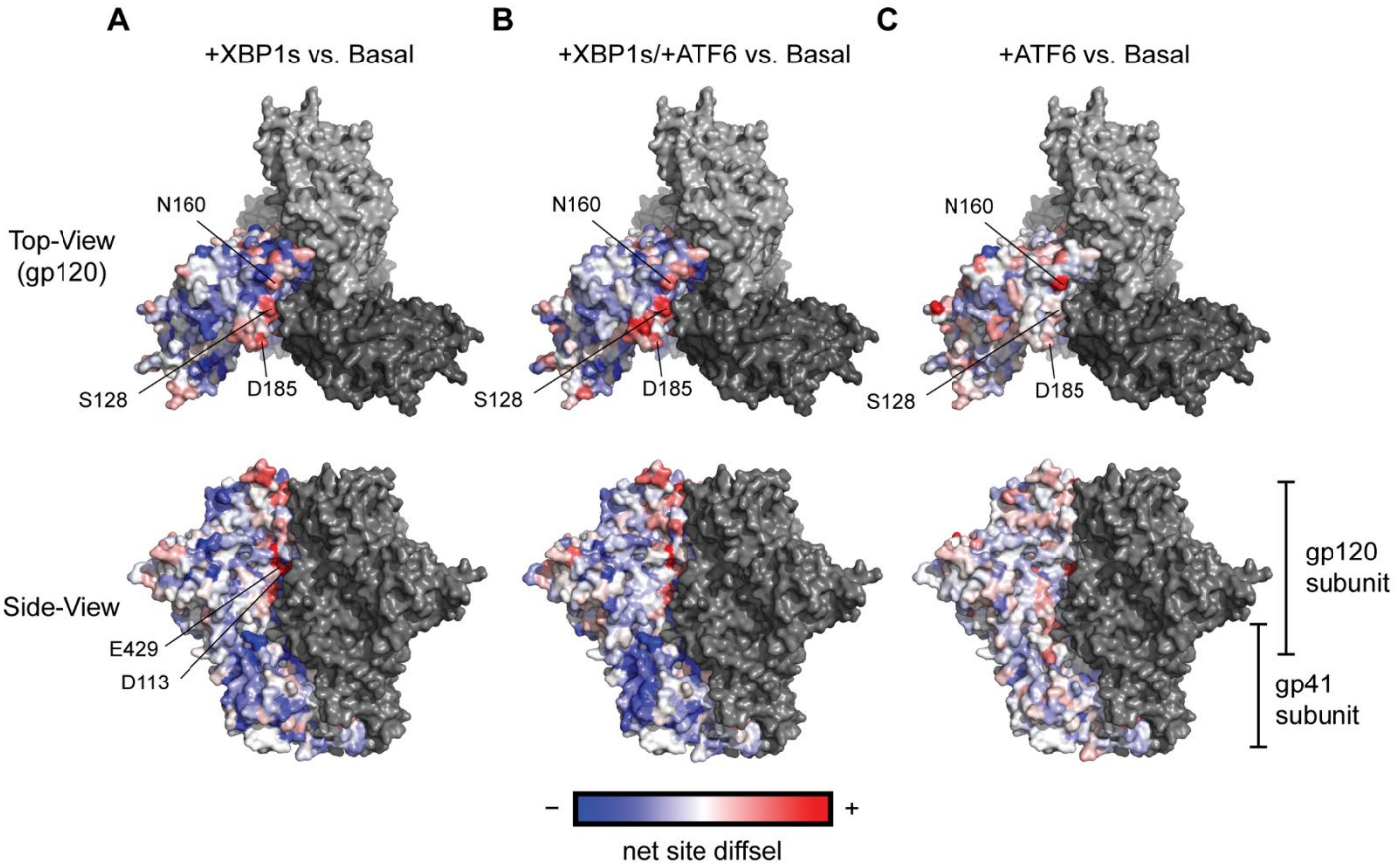
476 We next scrutinized Env regions directly involved in membrane fusion, since the principal function of
477 Env in the HIV replication cycle is to facilitate host cell entry via the fusion of viral and host membranes.
478 Briefly, upon binding to cell surface CD4 receptor and coreceptor, the fusion peptide in gp41 is inserted into the
479 cell membrane, and the two heptad repeat domains form a three-stranded coiled-coil that allows the anchoring
480 of Env to the host cell membrane (65). With the exception of CD4 contact sites, regions participating in mem-
481 brane fusion (**Fig 5**; 'Membrane fusion') experienced decreased mutational tolerance upon XBP1s induction. In
482 addition, the hydrophobic network of gp120 that undergoes conformational changes upon CD4 binding to trig-
483 ger membrane fusion (66) exhibited negative net site diffsel (**Fig 6C**).
484 Lastly, we focused further attention on regions of Env that may play important roles in Env folding and
485 stability. We observed a significant decrease in mutational tolerance for sites participating in the gp120–gp41
486 subunit contact (**Fig 5**; 'Subunit contact'). Next, we asked what the consequences of XBP1s induction are for
487 disulfide bonds and *N*-glycosylation sequons. Particularly noteworthy, we observed that every single cysteine
488 residue involved in disulfide bonds exhibited negative net site diffsel upon XBP1s induction (**Fig 5**; 'Disulfide
489 bond' and **Fig 6D**), consistent with the notion that the XBP1s-remodeled ER proteostasis environment strictly
490 quality controls disulfide bond formation in Env. The results were different for *N*-glycosylation sequons, even
491 though these residues can also promote ER protein folding and quality control by providing access to the ER's
492 lectin-based chaperone network (67). We observed an approximately equal number of sites in *N*-glycosylation
493 sequons that displayed positive and negative net site diffsel upon XBP1s induction (**Fig 5**; '*N*-glycosylation').
494 In fact, several *N*-glycosylation sequons displayed positive net site diffsel across all three enhanced ER proteo-
495 stasis environments (**Fig 6E**, **S13E Fig**, and **S13J Fig**). Among those *N*-glycosylation sequons displaying posi-
496 tive net site diffsel, all except N160 are highly variable (68). These observations add to the evidence that

497 mutational tolerance is more strongly constrained in conserved regions than in variable regions upon upregula-
498 tion of the host's ER proteostasis machinery.

499 The trends observed for the co-induction of XBP1s and ATF6 largely overlapped with XBP1s induction
500 only (**S10 Fig** and **S13A–E Fig**), except that CD4 contact sites exhibited a statistically significant decrease in
501 mutational tolerance whereas subunit contact sites did not. Consistent with the less striking reduction in muta-
502 tional tolerance observed upon ATF6 induction (**Fig 2C**), we observed that the impact of ATF6 induction was
503 minimal across Env sites when we assessed structural/functional groups independently (**S11 Fig** and **S13F–J**
504 **Fig**). Only the gp41 subunit exhibited a small, yet statistically significant decrease in mutational tolerance (**S11**
505 **Fig**; 'Subunits'), which agrees with our slope analysis of the sequence logo plots (**Fig 4B**–**4D**).

506 Finally, to evaluate structural regions whose mutational tolerance was particularly impacted by host ER
507 proteostasis network remodeling, we mapped net site diffsel values onto the Env crystal structure (**Fig 7**).
508 Whereas mutationally intolerant sites were distributed throughout the Env trimer, sites with enhanced muta-
509 tional tolerance upon XBP1s induction were located primarily at the apex of the Env trimer (**Fig 7A**). For in-
510 stance, N160, S128, and D185 were among the sites with the highest positive net site diffsel. Indeed, although
511 the magnitude of enhanced mutational tolerance varied, these sites exhibited positive net site diffsel in all host
512 ER proteostasis conditions tested. N160, S128, and D185 had similar net site diffsel values when XBP1s was
513 induced (**Fig 7A**) or when XBP1s and ATF6 were co-induced (**Fig 7B**), but N160 exhibited substantially higher
514 mutational tolerance when ATF6 was induced (**Fig 7C**). Notably, N160 belongs to the V2 apex, a well-charac-
515 terized epitope targeted by the broadly neutralizing antibodies PG9 (69), CH01 (70), CAP256.09 (71), and
516 PGT145 (72), and elimination of the N160 glycan was shown to confer antibody escape (37). In addition,
517 I165K, a fusion peptide inhibitor resistance mutation (70), was the single variant with the highest positive
518 diffsel when XBP1s and ATF6 were co-induced and the third highest positive diffsel when XBP1s was induced
519 alone, and was also confirmed in our immunoblots to not disrupt Env processing (**Fig 3F**). These observations
520 suggest that upregulation of host ER proteostasis factors, although generally constraining Env mutational

521 tolerance, can still strongly enhance mutational tolerance in regions of the Env protein in which adaptive muta-
522 tions are essential, including mutations at certain antibody- or drug-targeted regions of Env.




523

524 **Fig 7. Env sites with positive net site diffsel are clustered at the trimer apex.**
525 Average net site diffsel values across Env for (**A**) +XBP1s (**B**) +XBP1s/+ATF6, and (**C**) +ATF6, normalized to the basal
526 ER proteostasis environment, are mapped onto Env trimer crystal structure (PDBID 5FYK) (73). One monomer is colored
527 using net site diffsel as the color spectrum; negative net site diffsel residues are colored in blue, and positive net site diffsel
528 residues are colored in red. The remainder of the Env trimer is colored in grey. Diffsel values are provided in
529 https://github.com/yoon-jimin/2021_HIV_Env_DMS.

530 ## Discussion

531 Our results provide the first experimental evidence, to our knowledge, that UPR-mediated upregulation
532 of the ER proteostasis network can globally reduce the mutational tolerance of a client protein. The primary ER
533 proteostasis factors involved in driving this effect in Env are XBP1s-regulated, as the broad-scale effects of
534 ATF6 induction are more muted (**Fig 2C** and **2F**). This result agrees with our RNA-Seq data, where XBP1s in-
535 duction led to upregulation of a larger number of ER proteostasis factors, including those known to interact with
536 Env (24-28).

537 This observation is consistent with the impacts of cellular quality control factors on protein mutational
538 tolerance, where the available protein sequence space is restricted through degradation and reduced trafficking
539 of aberrantly folded protein variants (16-18). Previous studies established that Env is readily targeted to and de-
540 graded by ERAD (27, 28, 74), suggesting that destabilizing Env variants may be subjected to more rapid re-
541 moval by quality control factors in an enhanced ER proteostasis environment. Indeed, upon induction of
542 XBP1s, which upregulates many quality control and ERAD components, conserved regions of Env exhibit par-
543 ticularly large decreases in mutational tolerance (**Fig 5** and **S10 Fig**), where mutations are more likely to cause
544 protein misfolding. Rosetta $\Delta\Delta G$ predictions and immunoblotting experiments (**Fig 3**) confirmed that the vari-
545 ants with negative diffsel upon XBP1s induction were generally more destabilizing and exhibited larger pro-
546 cessing defects when compared to the positive diffsel variants.

547 While our evidence is consistent with the notion that UPR-regulated quality control factors are moderat-
548 ing Env mutational tolerance directly, some of the observed effects could also be secondary. For example, ER
549 proteostasis factors could post-translationally influence the folding or levels of endogenous proteins that regu-
550 late Env function or folding. Both of these phenomena are interesting. In addition, we note that the LAI strain of
551 HIV used in this study could have had lower mutational tolerance than HIV strains on average, as it was iso-
552 lated from a chronically infected individual and potentially accumulated a significant number of deleterious mu-
553 tations. In future studies, it will be interesting to examine effects in additional HIV strains.

554 This work augments the emerging evidence that host ER proteostasis machinery can fundamentally de-
555 fine the mutational tolerance of viral membrane proteins. Prior to this study, the consequences of ER proteosta-
556 sis network composition for the mutational tolerance of a membrane protein, whether viral or endogenous, had
557 only ever been investigated for one other protein – influenza hemagglutinin (32). We show that the host ER pro-
558 teostasis network also impacts Env mutational tolerance, implying the potential that this relationship is applica-
559 ble across multiple RNA viruses and diverse membrane proteins. Moreover, the present work reveals that the
560 interaction between host proteostasis and viral proteins is highly nuanced, and the outcome can differ for each
561 viral pathogen, which can be either due to intrinsic differences in the client protein or to differences in the cell
562 types the viruses infect. For example, hemagglutinin mutational tolerance is enhanced at febrile temperatures
563 upon XBP1s induction, with very minimal effects at a permissive temperature (32). Unlike hemagglutinin, the
564 majority of Env sites exhibited strongly decreased mutational tolerance upon upregulation of host ER proteosta-
565 sis factors, in this case even at a permissive temperature. Comparing our RNA-Seq data from SupT1$^{DAX}$ cells
566 with a previous characterization of HEK293$^{DAX}$ cells used in the hemagglutinin work (32), we observed that
567 73% (+XBP1s) and 58% (+XBP1s/+ATF6) of the transcripts upregulated in SupT1$^{DAX}$ cells were also upregu-
568 lated in the HEK293$^{DAX}$ cells (note that the +ATF6 condition was not tested in the previous study; **S11 Data**).
569 Differences in the UPR response in the two cell lines, as well structure and folding pathway differences in he-
570 magglutinin and Env themselves, may underpin the differing observations. Although beyond the scope of this
571 paper, a comparative analysis of the interactomes of hemagglutinin and Env with UPR-regulated ER proteosta-
572 sis factors, particularly focusing on the genes that were differentially enriched in HEK293$^{DAX}$ and SupT1$^{DAX}$
573 cells, may reveal specific contributors to this differing outcome.

574 Looking deeper into our observations for Env itself, this study highlights several Env regions that merit
575 further investigation with respect to their roles in Env folding and structure. For example, we found that sites
576 that constitute *N*-glycosylation sequons exhibited both positive and negative net site diffsel (**Fig 6E**, **S13E Fig**,
577 and **S13J Fig**). The fact that *N*-glycosylated residues were not particularly constrained may be a reflection of
578 the fact that they can act redundantly in endowing key interactions with lectin-based chaperone and quality

579 control pathways. Indeed, we previously showed that a nonnative *N*-glycosylation sequon can successfully ena-
580 ble calnexin/calreticulin-mediated ER client protein folding (67). In addition, while *N*-glycans in Env are im-
581 portant for antibody shielding and viral replication (75-78), there have been varying reports on whether the ma-
582 jority of the *N*-glycans are required for proper folding of Env (78, 79). The specific *N*-glycan sites that proved
583 particularly sensitive to XBP1s upregulation are likely to play some important role in the folding, quality con-
584 trol, and/or trafficking of Env. It will be interesting to explore the specific biophysical mechanisms underlying
585 our observations in future work.

586 Finally, we find that different sites within a single viral protein can respond differently to the selection
587 pressure imposed by the host ER proteostasis network (**Fig 5**, **Fig 6**, **S7 Fig**, **S8 Fig**, and **S13 Fig**). Contrary to
588 the global trend in decreased mutational tolerance, we observed many Env sites with positive net site diffsel,
589 especially at the trimer apex of Env (**Fig 7**). We discovered that N160, where a glycan is installed that is obliga-
590 tory for binding of the vast majority of V2 apex broadly neutralizing antibodies (80), exhibited enhanced muta-
591 tional tolerance in all three proteostasis environments and particularly when ATF6 was induced alone. We also
592 observed that I165K, an Env variant known to be fusion peptide inhibitor resistant (70), exhibited highly posi-
593 tive diffsel upon XBP1s induction. These observations indicate that, although the majority of Env sites exhib-
594 ited depletion of variants, important antibody- or drug-escape variants may be enriched upon upregulation of
595 host ER proteostasis network mechanisms. Thus, the host ER proteostasis environment can strongly influence
596 the mutational tolerance of specific Env variants that are of therapeutic interest.

597 In conclusion, our results establish that stress response-mediated upregulation of proteostasis networks
598 can actually restrict rather than increase accessible client protein sequence space, in contrast to most prior work
599 focused on the effects of individual chaperones. We also find that evolutionary interactions between viral pro-
600 teins and host proteostasis factors are specific to the virus type, as well as to specific regions of the viral protein.
601 We anticipate this knowledge will prove particularly valuable for ongoing efforts to target host proteostasis net-
602 work components for antiviral therapeutics (52, 81-86) and for the design of proteostasis network-targeted ther-
603 apeutic adjuvants that can prevent the emergence of viral variants that confer immune system escape or drug

604 resistance. More broadly, the principles observed here seem likely to prove generally applicable, not just to viral
605 proteins but also endogenous client proteins.

## Acknowledgements

The authors would like to thank Prof. Jesse Bloom (Fred Hutchinson Cancer Research Center) and Dr. Hugh Haddox (University of Washington) for providing Env plasmid libraries. We are also grateful for the support from the Tufts Technology Services and for the computing resources at the Tufts Research Cluster.

610 ## Materials and Methods

611 **Cell culture.** Human T lymphoblasts (SupT1 cells; ATCC) were grown in RPMI-1640 medium (Corning), sup-
612 plemented with 10% heat-inactivated fetal bovine serum (FBS, Cellgro), 1% penicillin/streptomycin/glutamine
613 (Cellgro) at 37 °C with 5% $CO_2$(g). TZM-bl reporter cells (NIH AIDS Research and Reference Reagent Pro-
614 gram; Cat. no. 1470) and HEK293T cells were cultured in DMEM (Corning) supplemented with 10% heat-inac-
615 tivated FBS, 1% penicillin/streptomycin/glutamine at 37 °C with 5% $CO_2$(g). Cell lines were periodically tested
616 for mycoplasma using the MycoSensor PCR Assay Kit (Agilent).

617 **Transfection of HEK293T cells with Env.** For transient expression of Env in HEK293T cells, we used the
618 Env gene from HIV-LAI in a pcDNA3.1 expression vector (Addgene) (87). Env variants were introduced by
619 site-directed mutagenesis (Agilent) and confirmed by Sanger sequencing of the Env gene (**S1 Table**).
620 HEK293T cells were plated in 6-well plates at a density of $7 \times 10^5$ cells/well and allowed to adhere overnight.
621 The next day, cells were transfected with 1.5 μg of eGFP in pcDNA3.1 (GFP control) or 0.15 μg eGFP and 1.35
622 μg of Env plasmid using Lipofectamine reagents (Thermo Fisher). After 16 h, the media was changed. After
623 another 24 h, cells were harvested for analysis.

624 **Plasmids to engineer SupT1$^{DAX}$ cells.** The following lentiviral destination vectors were used for stable cell line
625 construction: pLenti6/V5 Dest Gateway with a tetracycline repressor insert (Invitrogen) and blasticidin re-
626 sistance, pLenti CMV/TO Zeocin DEST with either human XBP1s insert (Addgene), and pLenti CMV hygro-
627 mycin DEST with a DHFR.ATF6(1-373) fusion, as previously described (41).

628 **Stable cell line engineering.** We generated a stable SupT1$^{DAX}$ cell line using a previously described method for
629 chemical genetic control of IRE1-XBP1s and ATF6 transcription factors (41). Specifically, SupT1 cells were
630 first transduced with lentivirus encoding a blasticidin-resistant tetracycline repressor and then with lentivirus
631 encoding zeocin resistant XBP1s. Transduction was performed by spinoculation with 2 μg/mL polybrene
632 (Sigma-Aldrich) at 1,240 × *g* for 1–1.5 h. Heterostable cell lines expressing the tetracycline repressor and
633 XBP1s were then selected using 10 μg/mL blasticidin (Gibco) and 50 μg/mg zeocin (Invitrogen). Single colony
634 lines were derived from the heterostable population by seeding 30–40 cells in a 96-well plate in 100 μl of RPMI

635 media without antibiotics for 10–14 days. Clonal populations were then selected and expanded in 24-well plates
636 in 500 μL of RPMI containing 10 μg/mL blasticidin and 50 μg/mL zeocin. Cells were grown to confluency and
637 then screened based on functional testing of the XBP1s construct using real-time polymerase chain reaction
638 (RT-PCR; described below) with or without 2 μg/mL doxycycline (dox; Alfa Aesar). The selected SupT1 single
639 colony cell line encoding tetracycline-inducible XBP1s was then transduced with lentivirus encoding
640 DHFR.ATF6(1–373) via the spinoculation protocol described above and stable cells were selected using 400
641 μg/mL hygromycin B (Gibco). The heterostable populations were then treated with vehicle, 2 μg/mL dox, 10
642 μM trimethoprim (TMP; Alfa Aesar), or 2 μg/mL dox and 10 μM trimethoprim and screened for function using
643 RT-PCR (described below) to obtain the final stably engineered SupT1$^{DAX}$ cell line.
644 **RT-PCR.** For RT-PCR of SupT1 cells to screen for stably engineered SupT1$^{DAX}$ cells with desired properties,
645 the SupT1 single colony cell line encoding tetracycline-inducible XBP1s was seeded at a density of $2 \times 10^5$
646 cells/well in a 6-well plate (Corning) in RPMI media and treated with 0.01 % DMSO or 2 μg/mL dox for 18 h.
647 SupT1$^{DAX}$ cells were seeded at a density of $2 \times 10^5$ cells/well in a 6-well plate in RPMI media and treated with
648 0.01 % DMSO, 2 μg/mL dox, 10 μM TMP, or 2 μg/mL dox and 10 μM TMP for 18 h. As a positive control for
649 UPR induction, the cells were treated with 10 μg/mL tunicamycin (Tm, Sigma-Aldrich) for 6 h. Cellular RNA
650 was harvested using the Omega RNA Extraction Kit with Homogenizer Columns (Omega Bio-tek). 1 μg RNA
651 was used to prepare cDNA using random primers (total reaction volume = 20 μL; Applied Biosystems High-
652 Capacity Reverse Transcription Kit). The reverse transcription reaction was diluted to 80 μL with water, and 2
653 μL of each sample was used for qPCR with 2 × Sybr Green (Roche) and primers for human *RPLP2* (housekeep-
654 ing gene), *HSPA5* (BIP), *HSP90B* (GRP94), *DNAJB9* (ERDJ4), and *SEC24D* (**S1 Table**). For qPCR data analy-
655 sis, all gene transcripts were normalized to that of *RPLP2*, and the fold-change in expression relative to DMSO-
656 treated cells was calculated.
657 For RT-PCR of HEK293T cells, cells were transfected with Env variants and cellular RNA was har-
658 vested using the Omega RNA Extraction Kit with Homogenizer Columns. As a positive control for UPR induc-
659 tion, GFP-transfected cells were treated with 2 μM thapsigargin (Tg; Sigma-Aldrich) for 6 h prior to RNA

660 harvest. The reverse transcription reaction was performed identically to SupT1 cells and 2 µL of each sample
661 was used for qPCR with 2 × Sybr Green and primers for human *RPLP2* (housekeeping gene), *SEC24D*, *HSPA5*
662 (BIP), *DNAJB9* (ERDJ4), and *HYOU1* (**S1 Table**). For qPCR data analysis, all gene transcripts were normal-
663 ized to that of *RPLP2*, and the fold-change in expression relative to GFP-transfected cells was calculated.
664 **RNA-Seq.** SupT1$^{DAX}$ cells were seeded in a 6-well plate at a density of 5 x $10^5$ cells/well in RPMI media in
665 quadruplicate. The cells were treated with 0.01 % DMSO (vehicle), 2 µg/mL dox (to activate the XBP1s tran-
666 scriptional response), 10 µM TMP (to activate the ATF6 transcriptional response), or 2 µg/mL dox and 10 µM
667 TMP (to simultaneously activate the XBP1s and ATF6 transcriptional responses) for 24 h. Cellular RNA was
668 harvested using the RNeasy Plus Mini Kit with QIAshredder homogenization columns (Qiagen). RNA-Seq li-
669 braries were prepared using the Kapa mRNA HyperPrep RNA-seq library construction kit (Kapa/Roche), with 6
670 min fragmentation at 94 °C and nine PCR cycles of final amplification and duplex barcoding. Libraries were
671 quantified using the Fragment Analyzer and qPCR before being sequenced on an Illumina HiSeq 2000 using
672 40-bp single-end reads in High Output mode.
673 Analyses were performed using previously described tools and methods (88). Reads were aligned
674 against hg19 (Feb., 2009) using bwa mem v. 0.7.12-r1039 [RRID:SCR_010910] with flags –t 16 –f, and map-
675 ping rates, fraction of multiply-mapping reads, number of unique 20-mers at the 5′ end of the reads, insert size
676 distributions and fraction of ribosomal RNAs were calculated using bedtools v. 2.25.0 [RRID:SCR_006646]
677 (89). In addition, each resulting bam file was randomly down-sampled to a million reads, which were aligned
678 against hg19, and read density across genomic features was estimated for RNA-Seq-specific quality control
679 metrics. For mapping and quantitation, reads were aligned against GRCh38/ENSEMBL 89 annotation using
680 STAR v. 2.5.3a with the following flags -runThreadN 8 –runMode alignReads –outFilter-Type BySJout –out-
681 FilterMultimapNmax 20 –alignSJoverhangMin 8 –alignSJDBoverhangMin 1 –outFilterMismatchNmax 999 –
682 alignIntronMin 10 –alignIntronMax 1000000 –alignMatesGapMax1000000 –outSAMtype BAM SortedBy-
683 Coordinate –quantMode TranscriptomeSAM with –genomeDir pointing to a 75nt-junction GRCh38 STAR suf-
684 fix array (90). Gene expression was quantitated using RSEM v. 1.3.0 [RRID:SCR_013027] with the following

685 flags for all libraries: rsem-calculate-expression –calc-pme –alignments -p 8 –forward-prob 0 against an annota-
686 tion matching the STAR SA reference (91). Posterior mean estimates (pme) of counts and estimated RPKM
687 were retrieved.

688 For differential expression analysis, dox-, TMP-, or dox and TMP-treated SupT1$^{DAX}$ cells were com-
689 pared against vehicle-treated SupT1$^{DAX}$ cells. Differential expression was analyzed in the R statistical environ-
690 ment (R v.3.4.0) using Bioconductor's DESeq2 package on the protein-coding genes only [RRID:SCR_000154]
691 (92). Dataset parameters were estimated using the estimateSizeFactors() and estimateDispersions() functions;
692 read counts across conditions were modeled based on a negative binomial distribution, and a Wald test was used
693 to test for differential expression (nbinomWaldtest(), all packaged into the DESeq() function), using the treat-
694 ment type as a contrast. Shrunken $\log_2$ fold-changes were calculated using the lfcShrink function. Fold-changes
695 and *p*-values were reported for each protein-coding gene. Upregulation was defined as a change in expression
696 level >1.5-fold relative to the basal environment with a non-adjusted *p*-value $<10^{-10}$. Gene ontology analyses
697 were performed using the online DAVID server, according to tools and methods presented by Huang *et al* (88).
698 The volcano plots were generated using EnhancedVolcano (**Fig 1B–D**; https://github.com/kevinblighe/En-
699 hancedVolcano).

700 **Gene Set Enrichment Analysis (GSEA).** Differential expression results from DESeq2 were retrieved, and the
701 "stat" column was used to pre-rank genes for GSEA analysis. These "stat" values reflect the Wald's test per-
702 formed on read counts as modeled by DESeq2 using the negative binomial distribution. Genes that were not ex-
703 pressed were excluded from the analysis. GSEA (desktop version, v3.0) (47, 93) was run in the pre-ranked
704 mode against MSigDB 7.0 C5 (Gene Ontology) set, using the official gene symbol as the key, with a weighted
705 scoring scheme, normalizing by meandiv, with 8958 gene sets retained, and 5000 permutations were run for *p*-
706 value estimation. Selected enrichment plots were visualized using a modified version of ReplotGSEA, in R
707 (https://github.com/PeeperLab/Rtoolbox/blob/master/R/ReplotGSEA.R).

708 **Resazurin metabolism assay.** SupT1$^{DAX}$ cells were seeded in 96-well plates (Corning) at a density of $1.5 \times 10^5$
709 cells/well in RPMI media and then treated with 0.1% DMSO, 2 μg/mL dox, 10 μM TMP, or 2 μg/mL dox and

710 10 μM TMP. 72 h post-treatment, 50 μL RPMI containing 0.025 mg/mL resazurin sodium salt (Sigma) was
711 added to the wells and mixed thoroughly. After 2 h of incubation, resorufin fluorescence (excitation 530 nm;
712 emission 590 nm) was quantified using a Take-3 plate reader (BioTeK). Experiments were conducted in biolog-
713 ical quadruplicate.
714 **HIV titering.** TZM-bl reporter cells were seeded at a density of $2.5 \times 10^4$ cells/well in 48-well plates. After 5 h,
715 the cells were infected with 100 μL of serially diluted infectious HIV viral inoculum containing 10 μg/ml
716 polybrene. Each sample was used to infect four technical replicates. After 48 h, the viral supernatant was re-
717 moved, and the cells were washed twice with PBS and then fixed with 4% paraformaldehyde (Thermo Scien-
718 tific) for 20 min. The fixed cells were washed twice with PBS and then stained with 4 mM potassium ferrocya-
719 nide, 4 mM ferricyanide, and 0.4 mg/mL 5-bromo-4-chloro-3-indolyl-*p*-D-galactopyranoside (X-Gal) in PBS at
720 37 °C for 50 min. The cells were washed with PBS, blue cells were counted manually under a microscope, and
721 infectious titers were calculated based on the number of blue cells per volume of viral inoculum.
722 **Deep mutational scanning.** Three biological replicate HIV libraries were generated from three previously pre-
723 pared independent Env mutant plasmid libraries (a generous gift from Prof. Jesse Bloom, University of Wash-
724 ington) following the previously reported protocol (22). Briefly, to generate the plasmid libraries, codon mutant
725 libraries of *env* were first created via PCR mutagenesis using codon tiling mutagenic primers ((55) For each co-
726 don except the starting methionine, N-terminal signal peptide, and the C-terminal cytoplasmic tail, primers with
727 a randomized NNN nucleotide triplet in the codon of interest were used to create the forward- and reverse-mu-
728 tagenesis primer pool, the two fragment PCR reactions were run, and the products were joined with additional
729 PCR reactions. The resulting *env* amplicons were cloned into a recipient plasmid that had *env* replaced by GFP,
730 and transformed into competent cells to prepare the plasmid library. For DMS, $SupT1^{DAX}$ cells were seeded in
731 T175 vented tissue culture flasks (Corning) at a density of $1.0 \times 10^8$ cells/flask in RPMI media. The cells were
732 pre-treated with 0.01% DMSO, 2 μg/mL dox, 10 μM TMP, or 2 μg/mL dox and 10 μM TMP for 18 h. Pre-
733 treated cells were infected with the p1viral libraries at a MOI of 0.005 based on the infectious (TZM-bl) titers.
734 In addition, one flask was either mock-infected (negative control) or infected with wild-type virus (to enable

735 error correction for DMS data analysis). To remove unbound virions from culture, 6 h post-infection the cells
736 were pelleted at 2,000 rpm for 5 min, washed twice with 25 mL PBS, and then resuspended in 50 mL of RPMI
737 media treated with 0.01% DMSO, 2 μg/mL dox, 10 μM TMP, or 2 μg/mL dox and 10 μM TMP. Cell pellets
738 were harvested 96 h post-infection by centrifuging the culture at 2,000 rpm for 5 min. Cell pellets were washed
739 twice with PBS and then resuspended in 1 mL of PBS. Aliquots (100 μL) were added to Eppendorf tubes and
740 stored at –80 °C for subsequent DNA extraction.

741 To generate samples for Illumina sequencing, non-integrated viral DNA was purified from aliquots of
742 frozen SupT1$^{DAX}$ cells using a mini-prep kit (Qiagen) and ~$10^7$ cells per prep. PCR amplicons of Env were pre-
743 pared from plasmid or mini-prepped non-integrated viral DNA by PCR following a previously described proto-
744 col (22). The amplicons were sequenced using barcoded-subamplicon sequencing, dividing Env into nine rather
745 than the previously reported six sub-amplicons. We note that it was necessary to exclude Env amino acid resi-
746 dues 31–34 from analysis because, after PCR optimization, we were unable to identify functional primers for
747 the first sub-amplicon that did not include these sites. As previously described, at least $10^6$ Env molecules were
748 PCR-amplified for preparation of sub-amplicon sequencing libraries to ensure sufficient sampling of viral li-
749 brary diversity (56). Briefly, this sequencing library preparation method appends unique, random barcodes and
750 part of the Illumina adapter to Env subamplicon molecules. In a second round of PCR, the complexity of the
751 uniquely barcoded subamplicons was controlled to be less than the sequencing depth, and the remainder of the
752 Illumina adapter was appended. The resulting libraries were sequenced on an Illumina HiSeq 2500 in rapid run
753 mode with 2 × 250 bp paired-end reads. The primers used are described in **S1 Table**.

754 **Deep mutational scanning data analysis.** The software dms_tools2 (https://jbloomlab.github.io/dms_tools2/)
755 (57) was used to align the deep-sequencing reads, count the number of times each codon mutation was observed
756 both before and after selection, calculate the diffsel for each Env variant, and generate sequence logo plots (**Fig**
757 **4A**, **S7 Fig**, and **S8 Fig**). The IPython notebook for code to perform this analysis is provided in
758 https://github.com/yoon-jimin/2021_HIV_Env_DMS. In sequence logo plots, regions with decreased muta-
759 tional tolerance were defined as regions of Env where there were more than 15 amino acid residues in a row

760 with slope < -1.5 (for +XBP1s and +XBP1s/ATF6) or < -1 (+ATF6). The slope at residue *i* was calculated using
761 the formula:

762 $$\frac{(\text{cumulative net site diffsel})_{i+5} - (\text{cumulative net site diffsel})_{i-5}}{10}$$

763 SAA was calculated via PDBePISA (**S9 Fig**) (94) using the crystal structure of BG505 SOSIP.664 (PDBID
764 5V8M) (95) and aligning to the LAI Env sequence. PDBePISA calculates the solvent-accessible surface area of
765 the monomer ('ASA') and the solvent-accessible surface area that is buried upon formation of interfaces ('Bur-
766 ied surface in interfaces'). 'Buried surface in interfaces' values were subtracted from 'ASA' values to obtain the
767 SAA of trimer. Ligands and antibodies were removed from the PDB file prior to SAA analysis. Site entropy
768 (Shannon entropy) was calculated using the Los Alamos HIV Sequence Database Shannon Entropy-One tool
769 (**S12 Fig**). The calculation was based on the consensus sequence generated from the 7590 HIV-1 Env sequences
770 in the Los Alamos HIV Sequence Database (one sequence per patient up to 2019). The net site diffsel values
771 were mapped onto Env crystal structure (PDBID 5FYK) (73) using PyMOL (**Fig 7**).

772 **Calculating changes in protein folding free energy upon mutation using Rosetta.** The cartesian_ddg appli-
773 cation in Rosetta version 3.13 was used to calculate $\Delta\Delta G$ of protein stability upon substitution (58). To prepare
774 the initial structure for the $\Delta\Delta G$ calculations, a homology model of the HIV-1 envelope protein for the LAI
775 strain was constructed using the Rosetta comparative modeling protocol, RosettaCM (96). Residues 31–664 of
776 the HIV Env protein from the HIV-1 JR-FL strain (PDB ID: 5FYK, chain G and chain B) were used as the tem-
777 plate structure (73, 97). The structure had a truncation at the membrane proximal external region of gp41, and
778 the homology model was constructed for the domains whose coordinate data were available. A Rosetta sym-
779 metry definition file was created using the make_symmdef_file application to prepare the HIV Env trimer struc-
780 ture (98). There were 34 residues whose coordinates were missing in chains G and B of PDB 5FYK, and in the
781 hybridization process, the missing residues in the threaded structure were patched using target sequence-based
782 fragments and ab initio folding (96). A total of 1,000 models were generated, and the lowest-energy HIV Env
783 trimer model that preserved the 10 disulfide bonds observed in the crystal structure was selected for $\Delta\Delta G$ calcu-
784 lations.

785 The HIV Env trimer structure was relaxed using the Rosetta FastRelax application, which performed
786 five cycles of sidechain repacking and energy minimization using the Rosetta energy function ref2015_cart (58,
787 99-101). A total of 20 relaxed decoys were generated, and the lowest energy structure was used as the input
788 wild-type structure for the cartesian_ddg calculation. In the cartesian_ddg calculation, the target residue was
789 substituted in all three chains of the trimer structure, and any neighboring residues within a 9-Å radius were re-
790 packed and energy minimized using the ref2015_cart energy function. This process was repeated five times to
791 produce five energy scores for the mutant and for the wild-type. The $\Delta\Delta G$ values were calculated by subtracting
792 the average wild-type scores from the average mutant scores. To better relate the predicted $\Delta\Delta G$ values to ex-
793 perimental values, the $\Delta\Delta G$ values were then scaled by a factor of 0.34, which was previously determined by
794 fitting $\Delta\Delta G$ values calculated using Rosetta to experimental $\Delta\Delta G$ values in units of kcal/mole (58). The result-
795 ing $\Delta\Delta G$ values were divided by three to obtain the predicted $\Delta\Delta G$ values for one monomer of the trimer.
796 **Immunoblots.** For immunoblotting of SupT1$^{DAX}$ cells for UPR target proteins, SupT1$^{DAX}$ cells were seeded in
797 T75 culture flasks in RPMI media and grown until cells attained a density of $1 \times 10^6$ cells/mL. Cells were then
798 treated with 0.01% DMSO (vehicle control), 2 μg/mL dox (+XBP1s), 10 μM TMP (+ATF6), or 2 μg/mL dox
799 and 10 μM TMP (+XBP1s/+ATF6) for 24 h. After treatment, cells were pelleted by centrifugation at 1,000 × g
800 for 5 min. Pellets were washed with 1 × PBS, and then lysed in radioimmunoprecipitation assay buffer (RIPA;
801 25 mM Tris pH 8.0, 0.5% (m/v) sodium deoxycholate, 150 mM NaCl, 0.1% (m/v) sodium dodecyl sulfate, 1%
802 (v/v) IGEPAL CA-630) and protease inhibitor tablet (Thermo Fisher). Lysates were cleared by centrifugation at
803 $2 \times 10^4 \times$ g for 20 min, and total protein concentration was quantified using bicinchoninic acid assay (Thermo
804 Fisher). 108 μg of total protein was analyzed for each sample. Blots were incubated with anti-BiP primary (Cell
805 Signaling Technology), anti-SEC24D primary (Abcam), and anti-β-actin primary (Sigma) and 680 RD and 800
806 CW secondary (LI-COR) antibodies, and imaged by scanning on an Odyssey infrared imager (LI-COR).
807 For immunoblotting of HEK293T cells, cells were transfected with Env variants, pelleted, washed with
808 1 × PBS, and then lysed in RIPA buffer and protease inhibitor tablet. Lysates were cleared by centrifugation at
809 $2 \times 10^4 \times$ g for 20 min, and total protein concentration was quantified using the bicinchoninic acid assay. 30 μg

810 of total protein was analyzed for each sample. Blots were incubated with anti-gp41 primary (ARP-13049; ob-
811 tained through the NIH HIV Reagent Program, contributed by Dr. George Lewis) and 680 RD secondary anti-
812 bodies, and imaged by scanning on an Odyssey infrared imager, followed by quantification using Image Studio.
813 **Statistical analyses.** Unless indicated otherwise, experiments were performed in biological triplicate with repli-
814 cates defined as independent experimental entireties (i.e., from plating the cells to acquiring the data). For deep
815 mutational scanning, each biological replicate mutant viral library was prepared from independently generated
816 mutant plasmid libraries, as previously reported (56). The mean of $\Delta\Delta G$ distributions (**Fig 3A** and **3B**) were
817 tested for significance using two-sample *t*-test in Graph Pad Prism. Densitometric analyses of immunoblots (**Fig**
818 **3E** and **3F**) were tested for statistical significance using one-way ANOVA in Graph Pad Prism, comparing the
819 mean of each variant to the mean of wild type. Diffsel values from deep mutational scanning were tested for
820 significance of deviation from zero (no relative enrichment or depletion), using a one-sample *t*-test in Graph
821 Pad Prism. For diffsel values and net site diffsel values, two-tailed *p*-values are reported to assess whether the
822 mean (net site) diffsel for enhanced ER proteostasis environments were significantly different from zero (**Fig**
823 **2C** and **2F**). For net site diffsel distributions for specific functional and structural groups, *p*-values were Bonfer-
824 roni-corrected for 20 tests (**Fig 5**, **S10 Fig**, and **S11 Fig**).

825 **References**

826 1. DePristo MA, Weinreich DM, Hartl DL. Missense meanderings in sequence space: A biophysical view
827 of protein evolution. Nat Rev Genet. 2005;6(9):678-87.
828 2. Wylie CS, Shakhnovich EI. A biophysical protein folding model accounts for most mutational fitness
829 effects in viruses. Proc Natl Acad Sci USA. 2011;108(24):9916-21.
830 3. Smith JM. Natural selection and the concept of a protein space. Nature. 1970;225(5232):563-4.
831 4. Ogbunugafor CB. A reflection on 50 years of John Maynard Smith's "Protein Space". Genetics.
832 2020;214(4):749-54.
833 5. Tokuriki N, Tawfik DS. Stability effects of mutations and protein evolvability. Curr Opin Struct Biol.
834 2009;19(5):596-604.
835 6. Guerrero RF, Scarpino SV, Rodrigues JV, Hartl DL, Ogbunugafor CB. Proteostasis environment shapes
836 higher-order epistasis operating on antibiotic resistance. Genetics. 2019;212(2):565-75.
837 7. Cowen LE, Lindquist S. Hsp90 potentiates the rapid evolution of new traits: Drug resistance in diverse
838 fungi. Science. 2005;309(5744):2185-9.
839 8. Queitsch C, Sangster TA, Lindquist S. Hsp90 as a capacitor of phenotypic variation. Nature.
840 2002;417(6889):618-24.
841 9. Rutherford SL, Lindquist S. Hsp90 as a capacitor for morphological evolution. Nature.
842 1998;396(6709):336-42.
843 10. Whitesell L, Santagata S, Mendillo ML, Lin NU, Proia DA, Lindquist S. Hsp90 empowers evolution of
844 resistance to hormonal therapy in human breast cancer models. Proc Natl Acad Sci USA. 2014;111(51):18297-
845 302.
846 11. Aguilar-Rodriguez J, Sabater-Munoz B, Montagud-Martinez R, Berlanga V, Alvarez-Ponce D, Wagner
847 A, et al. The molecular chaperone DnaK is a source of mutational robustness. Genome Biol Evol.
848 2016;8(9):2979-91.

849 12. Williams TA, Fares MA. The effect of chaperonin buffering on protein evolution. Genome Biol Evol.
850 2010;2:609-19.
851 13. Wyganowski KT, Kaltenbach M, Tokuriki N. GroEL/ES buffering and compensatory mutations
852 promote protein evolution by stabilizing folding intermediates. J Mol Biol. 2013;425(18):3403-14.
853 14. Cetinbas M, Shakhnovich EI. Catalysis of protein folding by chaperones accelerates evolutionary
854 dynamics in adapting cell populations. PLoS Comput Biol. 2013;9(11):e1003269.
855 15. Tokuriki N, Tawfik DS. Chaperonin overexpression promotes genetic variation and enzyme evolution.
856 Nature. 2009;459(7247):668-73.
857 16. Thompson S, Zhang Y, Ingle C, Reynolds KA, Kortemme T. Altered expression of a quality control
858 protease in E. coli reshapes the in vivo mutational landscape of a model enzyme. eLife. 2020;9:e53476.
859 17. Bershtein S, Mu W, Serohijos AW, Zhou J, Shakhnovich EI. Protein quality control acts on folding
860 intermediates to shape the effects of mutations on organismal fitness. Mol Cell. 2013;49(1):133-44.
861 18. Penn WD, McKee AG, Kuntz CP, Woods H, Nash V, Gruenhagen TC, et al. Probing biophysical
862 sequence constraints within the transmembrane domains of rhodopsin by deep mutational scanning. Sci Adv.
863 2020;6(10):eaay7505.
864 19. Sebastian RM, Shoulders MD. Chemical biology framework to illuminate proteostasis. Annu Rev
865 Biochem. 2020;89:529-55.
866 20. Walter P, Ron D. The unfolded protein response: from stress pathway to homeostatic regulation.
867 Science. 2011;334(6059):1081-6.
868 21. Cuevas JM, Geller R, Garijo R, López-Aldeguer J, Sanjuán R. Extremely high mutation rate of HIV-1 in
869 vivo. PLoS Biol. 2015;13(9):e1002251.
870 22. Haddox HK, Dingens AS, Bloom JD. Experimental estimation of the effects of all amino-acid mutations
871 to HIV's envelope protein on viral replication in cell culture. PLoS Pathog. 2016;12(12):e1006114.
872 23. Klein JS, Bjorkman PJ. Few and far between: how HIV may be evading antibody avidity. PLoS Pathog.
873 2010;6(5):e1000908.

874 24. Ou WJ, Bergeron JJ, Li Y, Kang CY, Thomas DY. Conformational changes induced in the endoplasmic
875 reticulum luminal domain of calnexin by Mg-ATP and Ca2+. J Biol Chem. 1995;270(30):18051-9.
876 25. Otteken A, Moss B. Calreticulin interacts with newly synthesized human immunodeficiency virus type 1
877 envelope glycoprotein, suggesting a chaperone function similar to that of calnexin. J Biol Chem.
878 1996;271(1):97-103.
879 26. Earl PL, Moss B, Doms RW. Folding, interaction with GRP78-BiP, assembly, and transport of the
880 human immunodeficiency virus type 1 envelope protein. J Virol. 1991;65(4):2047-55.
881 27. Zhou T, Frabutt DA, Moremen KW, Zheng YH. ERManI (Endoplasmic reticulum class I alpha-
882 mannosidase) is required for HIV-1 envelope glycoprotein degradation via endoplasmic reticulum-associated
883 protein degradation pathway. J Biol Chem. 2015;290(36):22184-92.
884 28. Casini A, Olivieri M, Vecchi L, Burrone OR, Cereseto A. Reduction of HIV-1 infectivity through
885 endoplasmic reticulum-associated degradation-mediated Env depletion. J Virol. 2015;89(5):2966-71.
886 29. Phillips AM, Gonzalez LO, Nekongo EE, Ponomarenko AI, McHugh SM, Butty VL, et al. Host
887 proteostasis modulates influenza evolution. eLife. 2017;6:e28652.
888 30. Phillips AM, Ponomarenko AI, Chen K, Ashenberg O, Miao J, McHugh SM, et al. Destabilized adaptive
889 influenza variants critical for innate immune system escape are potentiated by host chaperones. PLoS Biol.
890 2018;16(9):e3000008.
891 31. Geller R, Pechmann S, Acevedo A, Andino R, Frydman J. Hsp90 shapes protein and RNA evolution to
892 balance trade-offs between protein stability and aggregation. Nat Commun. 2018;9(1):1781.
893 32. Phillips AM, Doud MB, Gonzalez LO, Butty VL, Lin YS, Bloom JD, et al. Enhanced ER proteostasis
894 and temperature differentially impact the mutational tolerance of influenza hemagglutinin. eLife.
895 2018;7:e38795.
896 33. Aviner R, Frydman J. Proteostasis in viral infection: Unfolding the complex virus-chaperone interplay.
897 Cold Spring Harb Perspect Biol. 2020;12(3):a034090.

898 34. Starr TN, Greaney AJ, Addetia A, Hannon WW, Choudhary MC, Dingens AS, et al. Prospective
899 mapping of viral mutations that escape antibodies used to treat COVID-19. Science. 2021;371(6531):850-4.
900 35. Fulton BO, Sachs D, Beaty SM, Won ST, Lee B, Palese P, et al. Mutational analysis of measles virus
901 suggests constraints on antigenic variation of the glycoproteins. Cell Rep. 2015;11(9):1331-8.
902 36. Dingens AS, Haddox HK, Overbaugh J, Bloom JD. Comprehensive mapping of HIV-1 escape from a
903 broadly neutralizing antibody. Cell Host Microbe. 2017;21(6):777-87.e4.
904 37. Dingens AS, Arenz D, Weight H, Overbaugh J, Bloom JD. An antigenic atlas of HIV-1 escape from
905 broadly neutralizing antibodies distinguishes functional and structural epitopes. Immunity. 2019;50(2):520-
906 32.e3.
907 38. Dingens AS, Arenz D, Overbaugh J, Bloom JD. Massively parallel profiling of HIV-1 resistance to the
908 fusion inhibitor Enfuvirtide. Viruses. 2019;11(5):439.
909 39. Ashenberg O, Padmakumar J, Doud MB, Bloom JD. Deep mutational scanning identifies sites in
910 influenza nucleoprotein that affect viral inhibition by MxA. PLoS Pathog. 2017;13(3):e1006288.
911 40. Doud MB, Lee JM, Bloom JD. How single mutations affect viral escape from broad and narrow
912 antibodies to H1 influenza hemagglutinin. Nat Commun. 2018;9(1):1386.
913 41. Shoulders MD, Ryno LM, Genereux JC, Moresco JJ, Tu PG, Wu C, et al. Stress-independent activation
914 of XBP1s and/or ATF6 reveals three functionally diverse ER proteostasis environments. Cell Rep.
915 2013;3(4):1279-92.
916 42. Ryno LM, Wiseman RL, Kelly JW. Targeting unfolded protein response signaling pathways to
917 ameliorate protein misfolding diseases. Curr Opin Chem Biol. 2013;17(3):346-52.
918 43. Wong MY, DiChiara AS, Suen PH, Chen K, Doan ND, Shoulders MD. Adapting secretory proteostasis
919 and function through the unfolded protein response. Curr Top Microbiol Immunol. 2018;414:1-25.
920 44. Thielen BK, Klein KC, Walker LW, Rieck M, Buckner JH, Tomblingson GW, et al. T cells contain an
921 RNase-insensitive inhibitor of APOBEC3G deaminase activity. PLoS Pathog. 2007;3(9):1320-34.

922 45. Golumbeanu M, Desfarges S, Hernandez C, Quadroni M, Rato S, Mohammadi P, et al. Proteo-
923 transcriptomic dynamics of cellular response to HIV-1 infection. Sci Rep. 2019;9(1):213.
924 46. Navare AT, Sova P, Purdy DE, Weiss JM, Wolf-Yadlin A, Korth MJ, et al. Quantitative proteomic
925 analysis of HIV-1 infected CD4+ T cells reveals an early host response in important biological pathways:
926 protein synthesis, cell proliferation, and T-cell activation. Virology. 2012;429(1):37-46.
927 47. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set
928 enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl
929 Acad Sci USA. 2005;102(43):15545-50.
930 48. Grandjean JMD, Plate L, Morimoto RI, Bollong MJ, Powers ET, Wiseman RL. Deconvoluting stress-
931 responsive proteostasis signaling pathways for pharmacologic activation using targeted RNA sequencing. ACS
932 Chem Biol. 2019;14(4):784-95.
933 49. Yamamoto K, Yoshida H, Kokame K, Kaufman RJ, Mori K. Differential contributions of ATF6 and
934 XBP1 to the activation of endoplasmic reticulum stress-responsive cis-acting elements ERSE, UPRE and
935 ERSE-II. J Biochem. 2004;136(3):343-50.
936 50. Yamamoto K, Sato T, Matsui T, Sato M, Okada T, Yoshida H, et al. Transcriptional induction of
937 mammalian ER quality control proteins is mediated by single or combined action of ATF6alpha and XBP1. Dev
938 Cell. 2007;13(3):365-76.
939 51. Vidal RL, Sepulveda D, Tronscoso-Escudero P, Garcia-Huerta P, Gonzalez C, Plate L, et al. Enforced
940 dimerization between XBP1s and ATF6f enhances the protective effects of the UPR in models of
941 neurodegeneration. Mol Ther. 2021;29(5):1862-82.
942 52. Nekongo EE, Ponomarenko AI, Dewal MB, Butty VL, Browne EP, Shoulders MD. HSF1 activation can
943 restrict HIV replication. ACS Infect Dis. 2020;6(7):1659-66.
944 53. Wei X, Decker JM, Liu H, Zhang Z, Arani RB, Kilby JM, et al. Emergence of resistant human
945 immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. Antimicrob Agents
946 Chemother. 2002;46(6):1896-905.

947 54. Korber BT, Foley BT, Kuiken CL, Pillai SK, Sodroski JG. Numbering positions in HIV relative to
948 HXB2CG. Human Retroviruses and AIDS, 1998. p. 102-11.
949 55. Bloom JD. An experimentally determined evolutionary model dramatically improves phylogenetic fit.
950 Mol Biol Evol. 2014;31(8):1956-78.
951 56. Doud MB, Bloom JD. Accurate measurement of the effects of all amino-acid mutations on influenza
952 hemagglutinin. Viruses. 2016;8(6):155.
953 57. Bloom JD. Software for the analysis and visualization of deep mutational scanning data. BMC
954 Bioinformatics. 2015;16:168.
955 58. Park H, Bradley P, Greisen P, Jr., Liu Y, Mulligan VK, Kim DE, et al. Simultaneous optimization of
956 biomolecular energy functions on features from small molecules and macromolecules. J Chem Theory Comput.
957 2016;12(12):6201-12.
958 59. Day JR, Munk C, Guatelli JC. The membrane-proximal tyrosine-based sorting signal of human
959 immunodeficiency virus type 1 gp41 is required for optimal viral infectivity. J Virol. 2004;78(3):1069-79.
960 60. Groppelli E, Len AC, Granger LA, Jolly C. Retromer regulates HIV-1 envelope glycoprotein trafficking
961 and incorporation into virions. PLoS Pathog. 2014;10(10):e1004518.
962 61. Nakane S, Iwamoto A, Matsuda Z. The V4 and V5 variable loops of HIV-1 envelope glycoprotein are
963 tolerant to insertion of green fluorescent protein and are useful targets for labeling. J Biol Chem.
964 2015;290(24):15279-91.
965 62. Chiasson MA, Rollins NJ, Stephany JJ, Sitko KA, Matreyek KA, Verby M, et al. Multiplexed
966 measurement of variant abundance and activity reveals VKOR topology, active site and human variant impact.
967 eLife. 2020;9:e58026.
968 63. Jiang X, Burke V, Totrov M, Williams C, Cardozo T, Gorny MK, et al. Conserved structural elements in
969 the V3 crown of HIV-1 gp120. Nat Struct Mol Biol. 2010;17(8):955-61.
970 64. Bowder D, Hollingsead H, Durst K, Hu D, Wei W, Wiggins J, et al. Contribution of the gp120 V3 loop
971 to envelope glycoprotein trimer stability in primate immunodeficiency viruses. Virology. 2018;521:158-68.

972 65. Sáez-Cirión A, Arrondo JL, Gómara MJ, Lorizate M, Iloro I, Melikyan G, et al. Structural and
973 functional roles of HIV-1 gp41 pretransmembrane sequence segmentation. Biophys J. 2003;85(6):3769-80.
974 66. Ozorowski G, Pallesen J, de Val N, Lyumkis D, Cottrell CA, Torres JL, et al. Open and closed
975 structures reveal allostery and pliability in the HIV-1 envelope spike. Nature. 2017;547(7663):360-3.
976 67. Li RC, Wong MY, DiChiara AS, Hosseini AS, Shoulders MD. Collagen's enigmatic, highly conserved
977 N-glycan has an essential proteostatic function. Proc Natl Acad Sci USA. 2021;118(10):e2026608118.
978 68. Raska M, Novak J. Involvement of envelope-glycoprotein glycans in HIV-1 biology and infection. Arch
979 Immunol Ther Exp (Warsz). 2010;58(3):191-208.
980 69. Walker LM, Phogat SK, Chan-Hui PY, Wagner D, Phung P, Goss JL, et al. Broad and potent
981 neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. Science.
982 2009;326(5950):285-9.
983 70. Bonsignori M, Hwang KK, Chen X, Tsao CY, Morris L, Gray E, et al. Analysis of a clonal lineage of
984 HIV-1 envelope V2/V3 conformational epitope-specific broadly neutralizing antibodies and their inferred
985 unmutated common ancestors. J Virol. 2011;85(19):9998-10009.
986 71. Doria-Rose NA, Schramm CA, Gorman J, Moore PL, Bhiman JN, DeKosky BJ, et al. Developmental
987 pathway for potent V1V2-directed HIV-neutralizing antibodies. Nature. 2014;509(7498):55-62.
988 72. Walker LM, Huber M, Doores KJ, Falkowska E, Pejchal R, Julien JP, et al. Broad neutralization
989 coverage of HIV by multiple highly potent antibodies. Nature. 2011;477(7365):466-70.
990 73. Guillaume BE, Stewart-Jones GB, Soto C, Lemmin T, Chuang GY, Druz A, et al. Trimeric HIV-1-Env
991 structures define glycan shields from clades A, B, and G. Cell. 2016;165(4):813-26.
992 74. Miranda LR, Schaefer BC, Kupfer A, Hu Z, Franzusoff A. Cell surface expression of the HIV-1
993 envelope glycoproteins is directed from intracellular CTLA-4-containing regulated secretory granules. Proc
994 Natl Acad Sci USA. 2002;99(12):8031-6.

995 75. Nakayama EE, Shioda T, Tatsumi M, Xin X, Yu D, Ohgimoto S, et al. Importance of the N-glycan in
996 the V3 loop of HIV-1 envelope protein for CXCR-4- but not CCR-5-dependent fusion. FEBS Lett.
997 1998;426(3):367-72.
998 76. Lee WR, Syu WJ, Du B, Matsuda M, Tan S, Wolf A, et al. Nonrandom distribution of gp120 N-linked
999 glycosylation sites important for infectivity of human immunodeficiency virus type 1. Proc Natl Acad Sci USA.
000 1992;89(6):2213-7.
001 77. Quiñones-Kochs MI, Buonocore L, Rose JK. Role of N-linked glycans in a human immunodeficiency
002 virus envelope glycoprotein: effects on protein function and the neutralizing antibody response. J Virol.
003 2002;76(9):4199-211.
004 78. Li Y, Luo L, Rasool N, Kang CY. Glycosylation is necessary for the correct folding of human
005 immunodeficiency virus gp120 in CD4 binding. J Virol. 1993;67(1):584-8.
006 79. Rathore U, Saha P, Kesavardhana S, Kumar AA, Datta R, Devanarayanan S, et al. Glycosylation of the
007 core of the HIV-1 envelope subunit protein gp120 is not required for native trimer formation or viral infectivity.
008 J Biol Chem. 2017;292(24):10197-219.
009 80. Andrabi R, Voss JE, Liang CH, Briney B, McCoy LE, Wu CY, et al. Identification of common features
010 in prototype broadly neutralizing antibodies to HIV envelope V2 apex to facilitate vaccine design. Immunity.
011 2015;43(5):959-73.
012 81. Taguwa S, Yeh MT, Rainbolt TK, Nayak A, Shao H, Gestwicki JE, et al. Zika virus dependence on host
013 Hsp70 provides a protective strategy against infection and disease. Cell Rep. 2019;26(4):906-20 e3.
014 82. Taguwa S, Maringer K, Li X, Bernal-Rubio D, Rauch JN, Gestwicki JE, et al. Defining Hsp70
015 subnetworks in dengue virus replication reveals key vulnerability in flavivirus infection. Cell.
016 2015;163(5):1108-23.
017 83. Geller R, Vignuzzi M, Andino R, Frydman J. Evolutionary constraints on chaperone-mediated folding
018 provide an antiviral approach refractory to development of drug resistance. Genes Dev. 2007;21(2):195-205.

019 84. Almasy KM, Davies JP, Lisy SM, Tirgar R, Tran SC, Plate L. Small-molecule endoplasmic reticulum
020 proteostasis regulator acts as a broad-spectrum inhibitor of dengue and Zika virus infections. Proc Natl Acad
021 Sci USA. 2021;118(3):e2012209118.
022 85. Joshi P, Maidji E, Stoddart CA. Inhibition of heat shock protein 90 prevents HIV rebound. J Biol Chem.
023 2016;291(19):10332-46.
024 86. Heaton NS, Moshkina N, Fenouil R, Gardner TJ, Aguirre S, Shah PS, et al. Targeting viral proteostasis
025 limits influenza virus, HIV, and dengue virus infection. Immunity. 2016;44(1):46-58.
026 87. Joshi A, Sedano M, Beauchamp B, Punke EB, Mulla ZD, Meza A, et al. HIV-1 Env glycoprotein
027 phenotype along with immune activation determines CD4 T cell loss in HIV patients. J Immunol.
028 2016;196(4):1768-79.
029 88. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: Paths toward the
030 comprehensive functional analysis of large gene lists. Nucleic Acids Research. 2009;37(1):1-13.
031 89. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features.
032 Bioinformatics. 2010;26(6):841-2.
033 90. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-
034 seq aligner. Bioinformatics. 2013;29(1):15-21.
035 91. Li B, Dewey CN. RSEM: Accurate transcript quantification from RNA-Seq data with or without a
036 reference genome. BMC Bioinformatics. 2011;12:323.
037 92. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data
038 with DESeq2. Genome Biol. 2014;15(12):550.
039 93. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-
040 responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat
041 Genet. 2003;34(3):267-73.
042 94. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol.
043 2007;372(3):774-97.

044 95. Lee JH, Andrabi R, Su CY, Yasmeen A, Julien JP, Kong L, et al. A broadly neutralizing antibody
045 targets the dynamic HIV envelope trimer apex via a long, rigidified, and anionic β-hairpin structure. Immunity.
046 2017;46(4):690-702.
047 96. Song Y, DiMaio F, Wang RY, Kim D, Miles C, Brunette T, et al. High-resolution comparative modeling
048 with RosettaCM. Structure. 2013;21(10):1735-42.
049 97. Zhang P, Kwon AL, Guzzo C, Liu Q, Schmeisser H, Miao H, et al. Functional anatomy of the trimer
050 apex reveals key hydrophobic constraints that maintain the HIV-1 envelope spike in a closed state. mBio.
051 2021;12(2):e00090-21.
052 98. DiMaio F, Leaver-Fay A, Bradley P, Baker D, Andre I. Modeling symmetric macromolecular structures
053 in Rosetta3. PLoS One. 2011;6(6):e20450.
054 99. Alford RF, Leaver-Fay A, Jeliazkov JR, O'Meara MJ, DiMaio FP, Park H, et al. The Rosetta all-atom
055 energy function for macromolecular modeling and design. J Chem Theory Comput. 2017;13(6):3031-48.
056 100. Maguire JB, Haddox HK, Strickland D, Halabiya SF, Coventry B, Griffin JR, et al. Perturbing the
057 energy landscape for improved packing during computational protein design. Proteins. 2021;89(4):436-49.
058 101. Khatib F, Cooper S, Tyka MD, Xu K, Makedon I, Popovic Z, et al. Algorithm discovery by protein
059 folding game players. Proc Natl Acad Sci USA. 2011;108(47):18949-53.
060

061 ## Supporting Information Captions

062 **S1 Table Primers for Env sequencing, RT-PCR, and site-directed mutagenesis**

063 **S2 Table Complete citations for structural and functional groups**

064 **S1 Data Complete RNA-Seq differential expression analysis**

065 **S2 Data Complete GSEA**

066 **S3 Data Resazurin assay and TZM-bl assay**

067 **S4 Data Library coverage**

068 **S5 Data Complete ΔΔ*G* analysis data**

069 **S6 Data Immunoblot densiometric analysis**

070 **S7 Data RT-PCR of UPR genes upon transfection of Env variants**

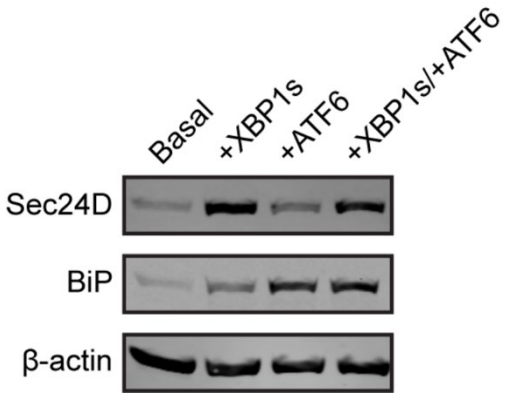
071 **S8 Data Cumulative net site diffsel**

072 **S9 Data Surface accessible area**

073 **S10 Data Site entropy**

074 **S11 Data Transcriptome comparison of HEK293$^{DAX}$ cells and SupT1$^{DAX}$ cells**

075




076

077 **S1 Fig. Immunoblot of SupT1$^{DAX}$ cells shows that the XBP1s and ATF6 pathways are successfully and**

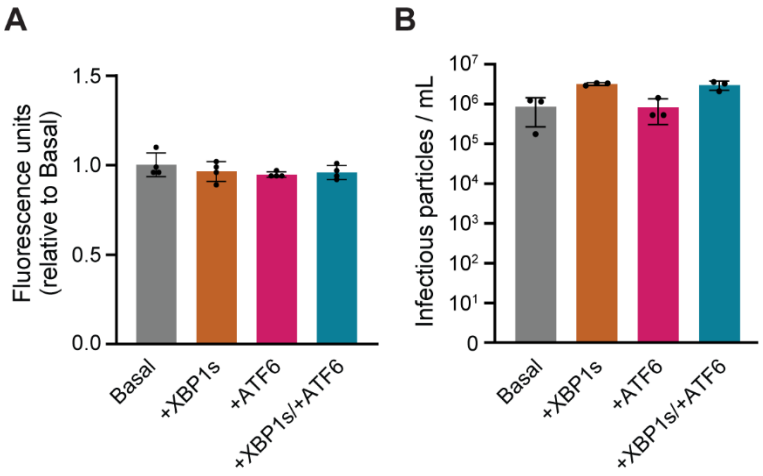
078 **differentially induced.** Representative immunoblot image showing specific upregulation of XBP1s (Sec24D)

079 and ATF6 (BiP) protein targets in SupT1$^{DAX}$ cells upon vehicle treatment (Basal), dox treatment (+XBP1s),

080 TMP treatment (+ATF6), and co-treatment of dox and TMP (+XBP1s/+ATF6).

081

082




083

084 **S2 Fig. ER proteostasis perturbation has no deleterious effects on cell viability and does not restrict HIV**

085 **replication.**

086 (**A**) Induction of XBP1s, ATF6, or co-induction of XBP1s and ATF6 did not alter metabolic activity of SupT1

087 cells, as measured by a resazurin assay. The average of biological quadruplicates is plotted with error bars rep-

088 resenting the standard deviation. Individual data points are also shown. (**B**) Induction of XBP1s and co-induc-

089 tion of XBP1s and ATF6 did not restrict and actually slightly increased HIV infectious titers, while induction of

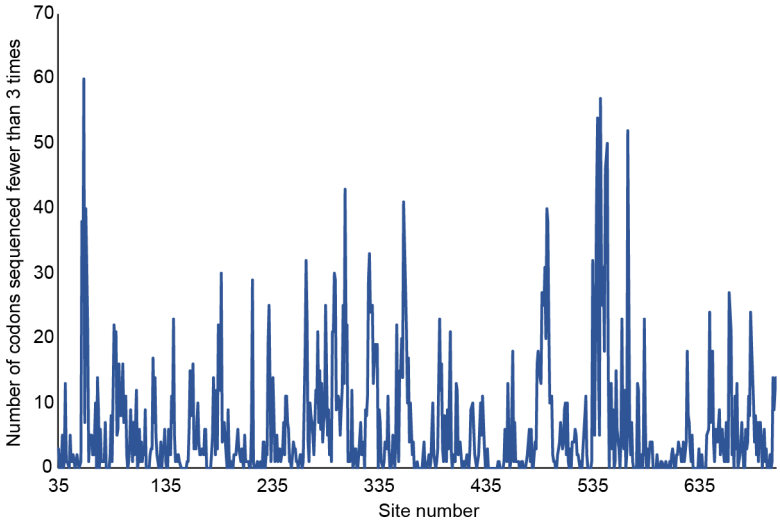
090 ATF6 did not influence HIV replication in SupT1 cells, as measured by TZM-bl infectious units. The average

091 of biological triplicates is plotted with error bars representing the standard deviation. Individual data points are

092 also shown. For **A** and **B**, replicate data are provided in **S3 Data**.

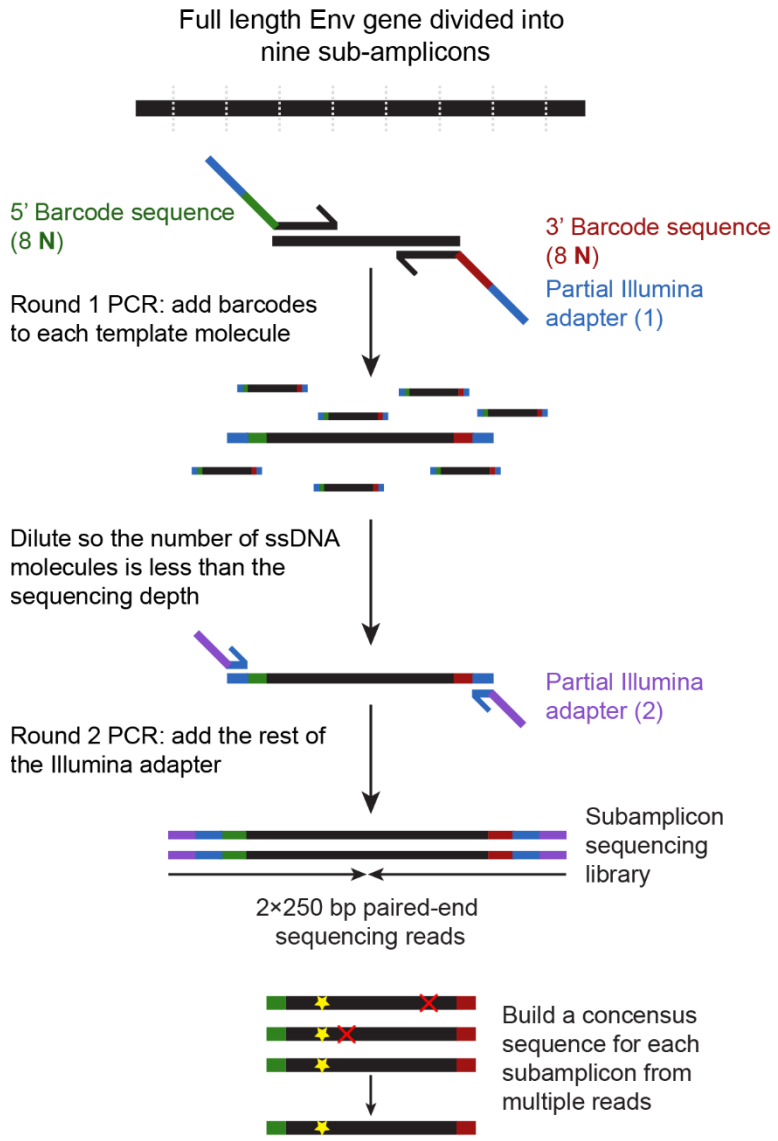
093

094

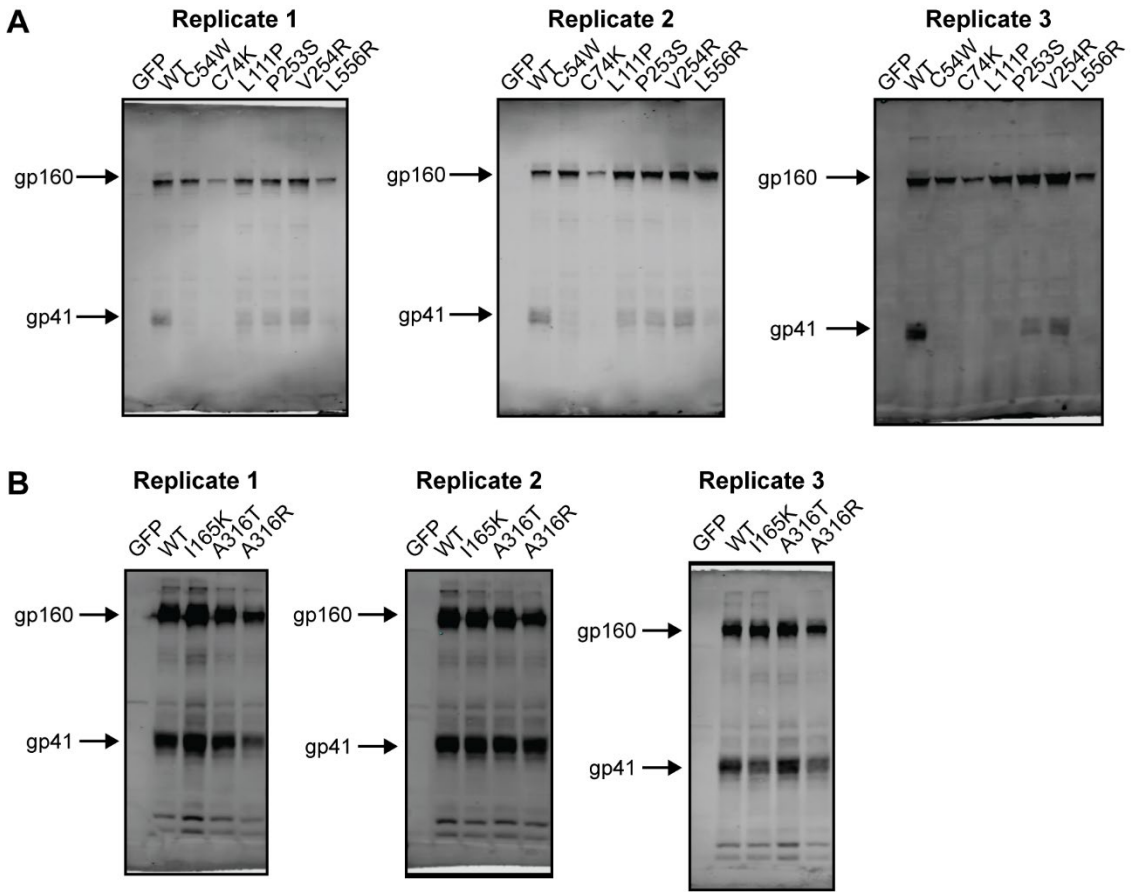



095

096 **S3 Fig. Library coverage was generally consistent throughout the Env sequence.** The number of codons
097 observed fewer than three times after summing the codon counts over the three biological replicate libraries is
098 plotted against the amino-acid site number. Sites with lower coverage were not localized to any specific domain
099 of structural or functional importance. Data values for library coverage are provided in **S4 Data**.

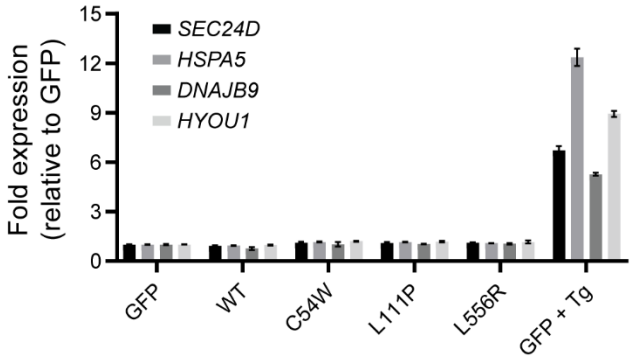



**S4 Fig. Subamplicon sequencing strategy ensures greater accuracy of reads during deep sequencing.** The full length Env gene was divided into nine subamplicons. In the first round of PCR, unique, random barcodes and part of the Illumina adapter were appended to the Env subamplicon molecules. In a second 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 was appended. The resulting libraries were sequenced on an Illumina HiSeq 2500 in rapid run mode with 2 × 250 bp paired-end reads.



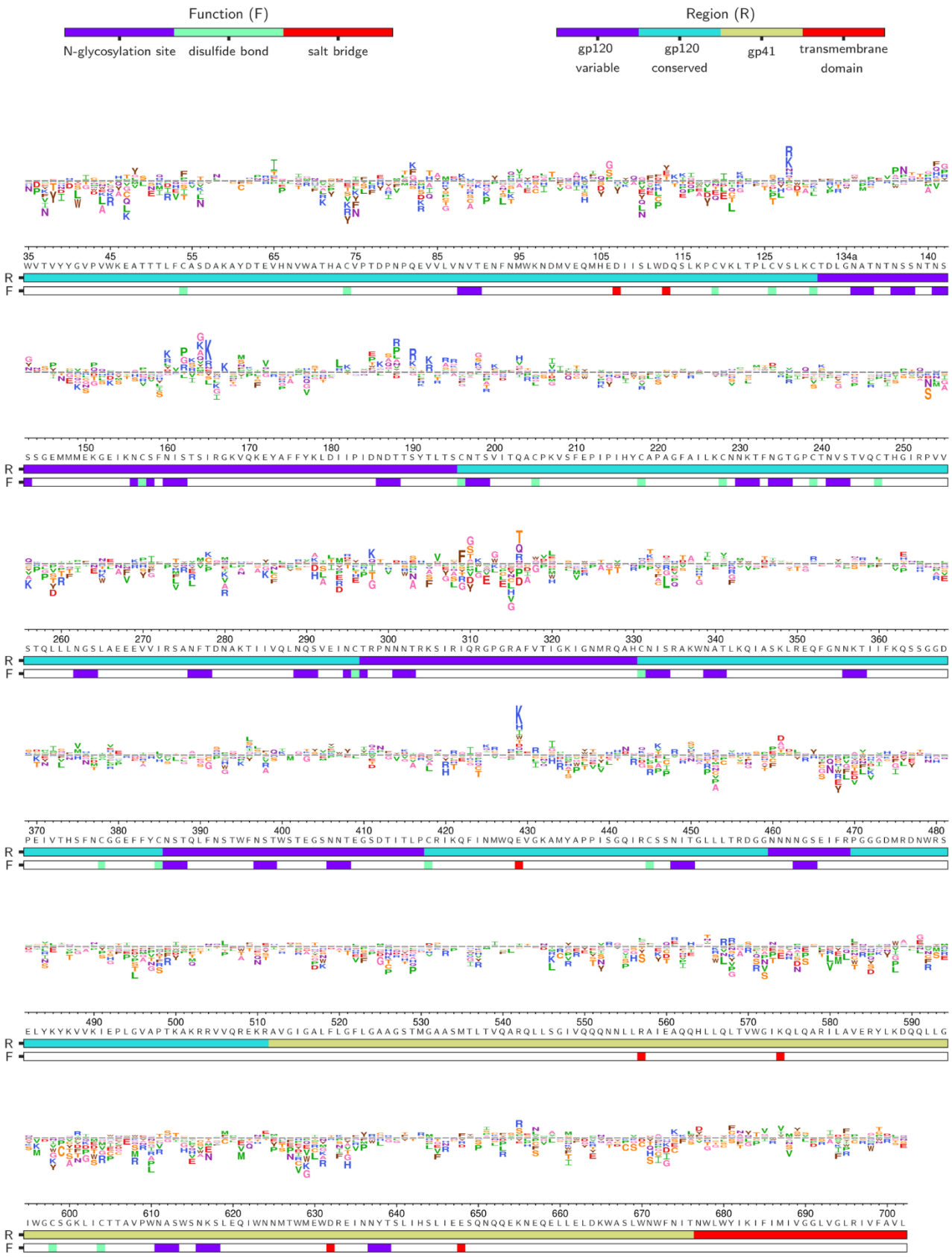

**S5 Fig. Env variants with negative diffsel exhibit processing defects.** Immunoblots in biological triplicates showing gp160 and gp41 bands for selected variants with (**A**) negative diffsel and (**B**) positive diffsel upon XBP1s induction.

112



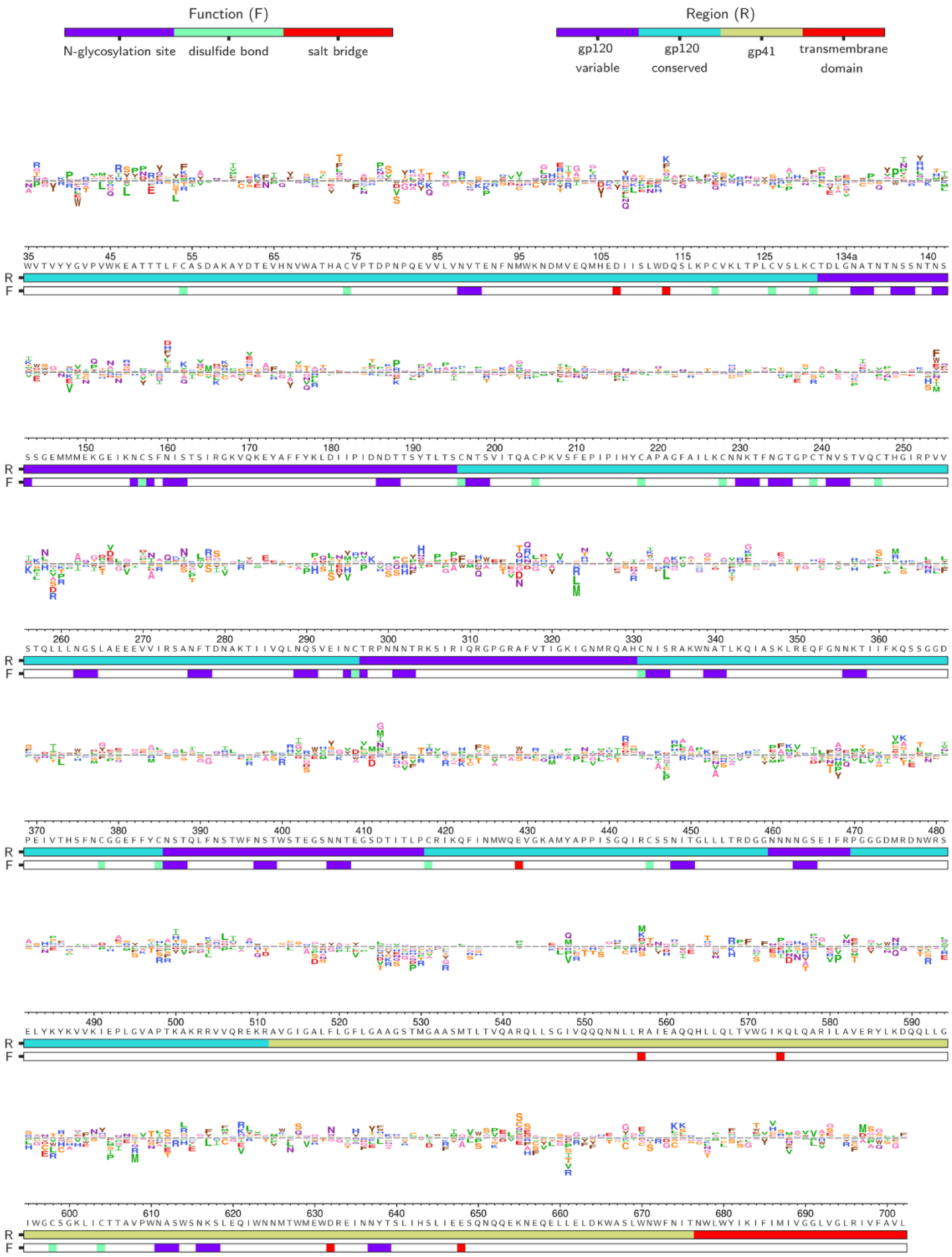

113

114 **S6 Fig. Transient transfection of Env variants with highly negative diffsel do not induce UPR.** RT-PCR
115 analysis of *SEC24D*, *HSPA5*, *DNAJB9*, and *HYOU1* in HEK293T cells expressing GFP (negative control),
116 wild-type Env, and three Env variants that were strongly negatively selected in +XBP1s vs. Basal (C54W,
117 L111P, L556R). As a positive control for UPR induction, HEK293T cells expressing GFP were treated with
118 thapsigargin (Tg; 2 μM) for 6h (GFP + Tg). RT-PCR data are presented as fold-increase relative to GFP-trans-
119 fected negative control. RT-PCR data values are provided in **S7 Data**.



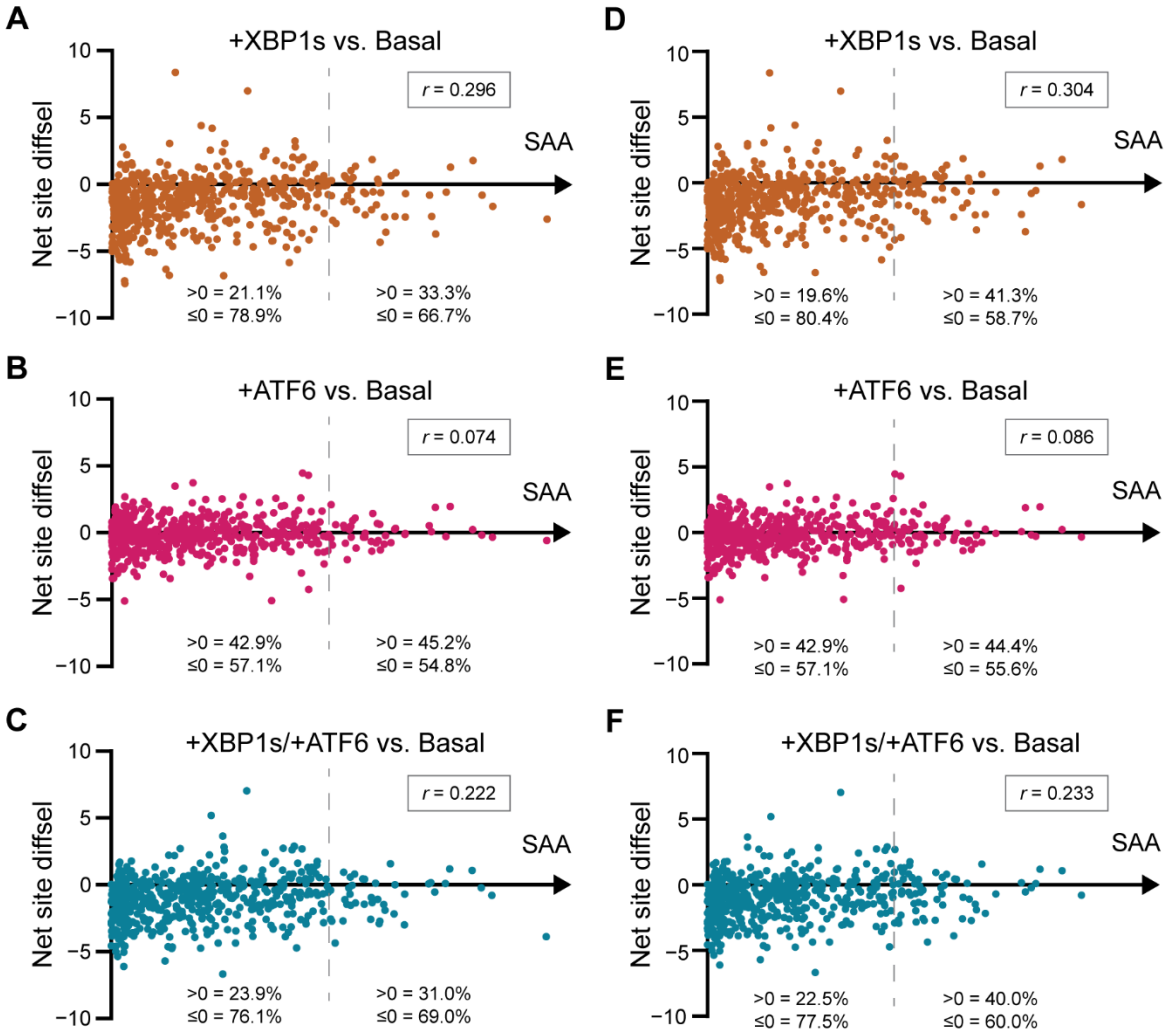

**S7 Fig. Sequence logo plots reveal diffsel across Env upon co-induction of XBP1s and ATF6.**

Logo plot displaying averaged diffsel for +XBP1s/+ATF6 normalized to the basal proteostasis environment. The height of the amino-acid abbreviation corresponds to the magnitude of diffsel. The amino-acid abbreviations are colored based on the side-chain properties: negatively charged (D, E; red), positively charged (H, K R; blue), polar uncharged (C, S, T; orange / N, Q; purple), small nonpolar (A, G; pink), aliphatic (I, L, M, P, V; green), and aromatic (F, W, Y; brown). The numbers and letters below the logos indicate the Env site in HXB2 numbering and the identity of the wild-type amino acid for that site, respectively. The color bar below the logos indicates the function (F) that the site is involved in (*N*-glycosylation site (purple), disulfide bond (green), or salt bridge (red)) or the region (R) of Env that the site belongs to (gp120-variable (purple), gp120-conserved (cyan), gp41 (yellow), or transmembrane domain (red); the sites that belong to the five variable loops of gp120 were categorized as 'gp120-variable', and the sites that are not included in the five variable loops were categorized as 'gp120-conserved'). Only variants that were present in all three pre-selection viral libraries and exhibited diffsel in the same direction across all three biological triplicates are plotted here. Diffsel values as well as unfiltered logo plots for each individual replicate are provided in https://github.com/yoon-jimin/2021_HIV_Env_DMS.



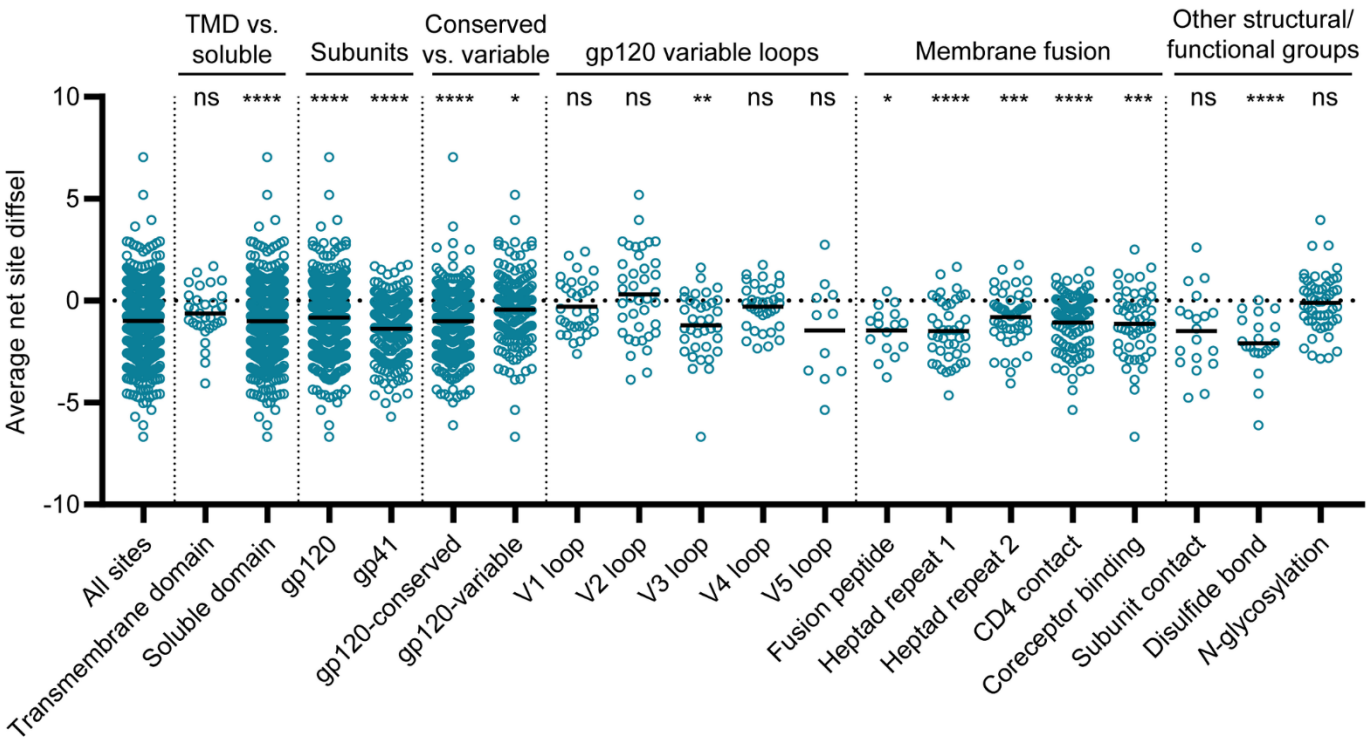

**S8 Fig. Sequence logo plots reveal diffsel across Env upon induction of ATF6.**

Logo plot displaying averaged diffsel for +ATF6 normalized to the basal proteostasis environment. The height of the amino-acid abbreviation corresponds to the magnitude of diffsel. The amino-acid abbreviations are colored based on the side-chain properties: negatively charged (D, E; red), positively charged (H, K R; blue), polar uncharged (C, S, T; orange / N, Q; purple), small nonpolar (A, G; pink), aliphatic (I, L, M, P, V; green), and aromatic (F, W, Y; brown). The numbers and letters below the logos indicate the Env site in HXB2 numbering and the identity of the wild-type amino acid for that site, respectively. The color bar below the logos indicates the function (F) that the site is involved in (*N*-glycosylation site (purple), disulfide bond (green), or salt bridge (red)) or the region (R) of Env that the site belongs to (gp120-variable (purple), gp120-conserved (cyan), gp41 (yellow), or transmembrane domain (red); the sites that belong to the five variable loops of gp120 were categorized as 'gp120-variable', and the sites that are not included in the five variable loops were categorized as 'gp120-conserved'). Only variants that were present in all three pre-selection viral libraries and exhibited diffsel in the same direction across all three biological triplicates are plotted here. Diffsel values as well as unfiltered logo plots for each individual replicate are provided in https://github.com/yoon-jimin/2021_HIV_Env_DMS.



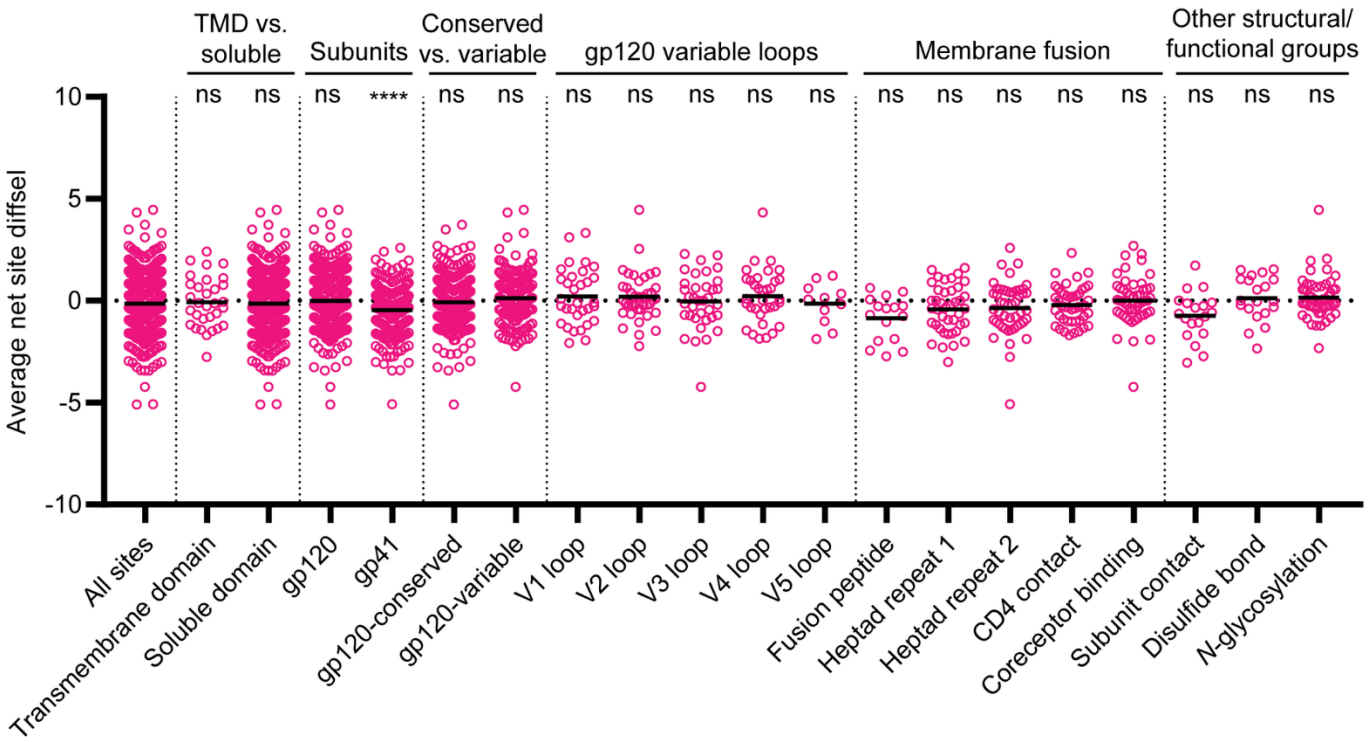

**S9 Fig. Env net site diffsel is not correlated with surface accessible area (SAA).**

Average net site diffsel values plotted against the SAA of Env monomer (**A–C**) and trimer (**D–F**). Average net site diffsel values for (**A**, **D**) +XBP1s, (**B**, **E**) +ATF6, and (**C**, **F**) +XBP1s/+ATF6 were normalized to the basal ER proteostasis environment and plotted against the SAA at each site. The percentages of variants with positive and negative net site diffsel for the left and right half of the plot are stated, as well as the Pearson correlation coefficient *r*. SAA was calculated using PDBePISA (94) with PDBID 5V8M (95), where SAA = 0 corresponds to a buried site. SAA data values are provided in **S9 Data**.

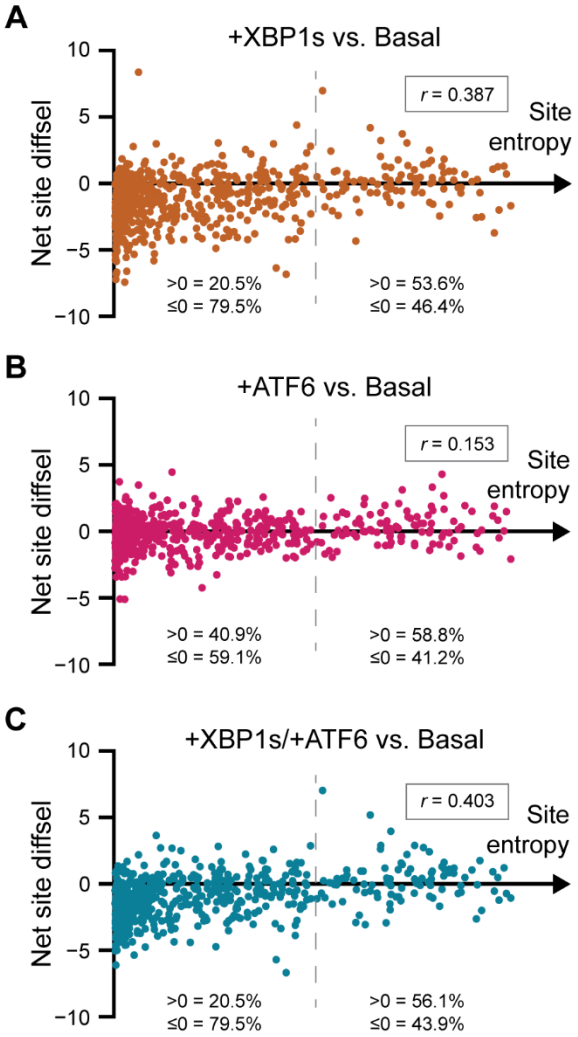



**S10 Fig. Impact of combined induction of XBP1s and ATF6 on mutational tolerance varies across Env structural elements.**

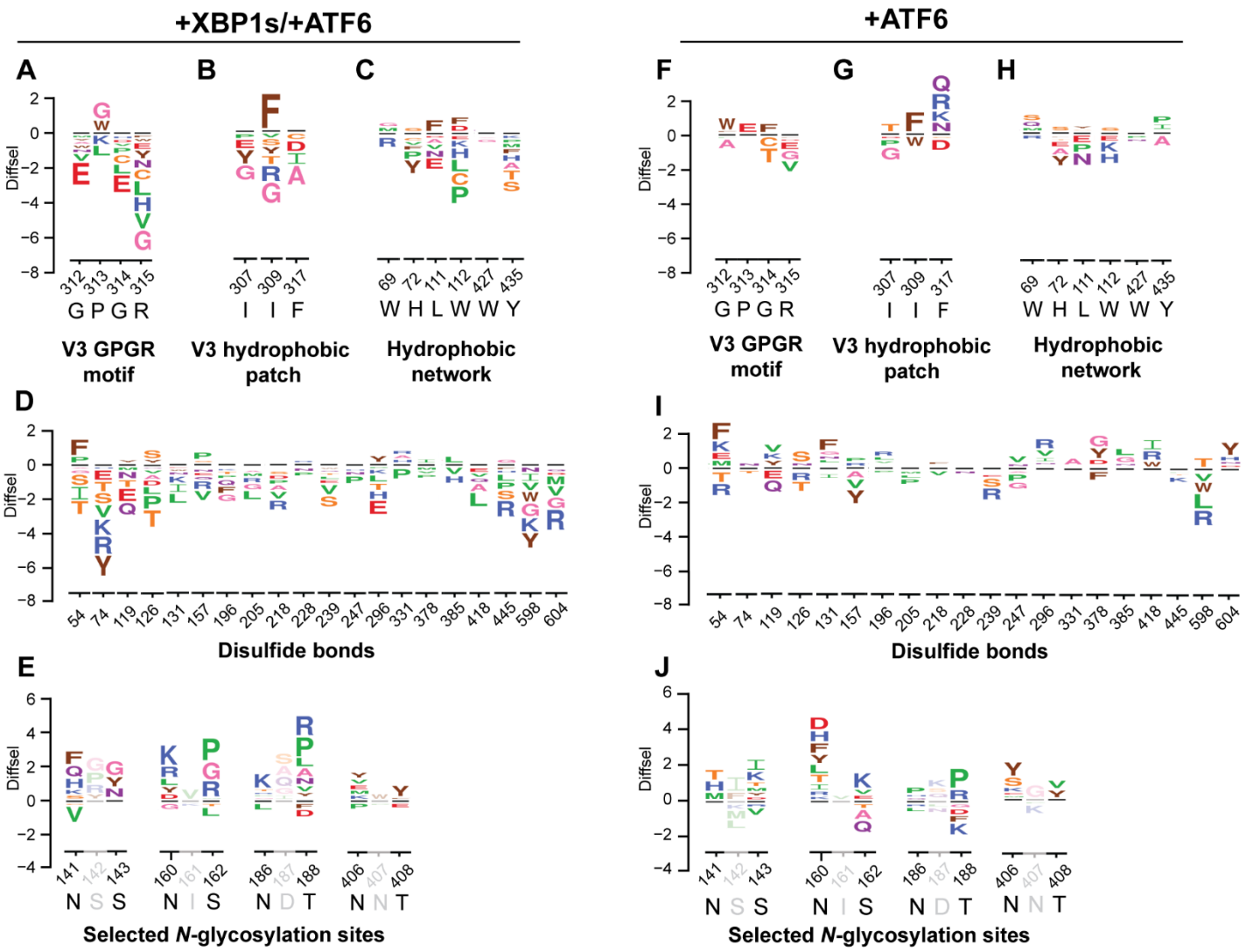
Average net site diffsel for the +XBP1s/+ATF6 ER proteostasis environment normalized to the basal ER proteostasis environment, where the means of distributions are indicated by black horizontal lines. Sites are sorted by TMD vs. soluble, subunits, conserved vs. variable regions of gp120, five variable loops of gp120, regions important for membrane fusion, and other structural/functional groups. For 'TMD vs. soluble', all sites that do not belong to the TMD were categorized as 'soluble'. For 'Conserved vs. variable', the sites that belong to the five variable loops of gp120 were categorized as 'gp120-variable', and the sites that are not included in the five variable loops were categorized as 'gp120-conserved'. Significance of deviation from null (net site diffsel = 0, no selection) was tested using a one-sample *t*-test. The derived *p*-values were Bonferroni-corrected for 20 tests and *, **, ***, and **** represent adjusted two-tailed *p*-values of <0.05, <0,01, <0.001, and <0.0001, respectively. Diffsel values are provided in https://github.com/yoon-jimin/2021_HIV_Env_DMS. Assignments for these structural regions are provided in **S2 Table**.




**S11 Fig. Impact of ATF6 induction on mutational tolerance varies across Env structural elements.** Average net site diffsel for the +ATF6 ER proteostasis environment normalized to the basal ER proteostasis environment, where the means of distributions are indicated by black horizontal lines. Sites are sorted by TMD vs. soluble, subunits, conserved vs. variable regions of gp120, five variable loops of gp120, regions important for membrane fusion, and other structural/functional groups. For 'TMD vs. soluble', all sites that do not belong to the TMD were categorized as 'soluble'. For 'Conserved vs. variable', the sites that belong to the five variable loops of gp120 were categorized as 'gp120-variable', and the sites that are not included in the five variable loops were categorized as 'gp120-conserved'. Significance of deviation from null (net site diffsel = 0, no selection) was tested using a one-sample *t*-test. The derived *p*-values were Bonferroni-corrected for 20 tests and *, **, ***, and **** represent adjusted two-tailed *p*-values of <0.05, <0,01, <0.001, and <0.0001, respectively. Diffsel values are provided in https://github.com/yoon-jimin/2021_HIV_Env_DMS. Assignments for these structural regions are provided in **S2 Table**.




**S12 Fig. Enhanced mutational tolerance is observed more frequently at sites with high site entropy.** Average net site diffsel values across Env for (**A**) +XBP1s (**B**) +ATF6, and (**C**) +XBP1s/+ATF6 are normalized to the basal ER proteostasis environment and plotted against the site entropy at each site. The percentages of variants with positive and negative net site diffsel for the left and right half of the plot are stated, as well as the Pearson correlation coefficient *r*. Site entropy data values are provided in **S10 Data**.




**S13 Fig. Diverse functional elements of Env respond differently to combined induction of XBP1s and ATF6, and ATF6 induction.**

Selected sequence logo plots for the +XBP1s/+ATF6 (**A–E**) and +ATF6 (**F–J**) ER proteostasis environments normalized to the basal ER proteostasis environment for (**A**, **F**) the conserved GPGR motif of the V3 loop, (**B**, **G**) the hydrophobic patch of the V3 loop, (**C**, **H**) the hydrophobic network of gp120 important for CD4 binding, (**D**, **I**) cysteine residues participating in disulfide bonds, and (**E**, **J**) selected *N*-glycosylation sequons (N-X-S/T) that exhibited positive net site diffsel in all three remodeled proteostasis environments. The height of the amino-acid abbreviation corresponds to the magnitude of diffsel. The numbers and letters below the logos indicate the Env site in HXB2 numbering and the wild-type amino acid for that site, respectively. Only variants that were present in all three pre-selection viral libraries and exhibited diffsel in the same direction across the biological

203 triplicates are plotted. All logo plots were generated on the same scale. Diffsel values are provided in

204 https://github.com/yoon-jimin/2021_HIV_Env_DMS. Assignments for these functional regions are provided in

205 **S2 Table**.