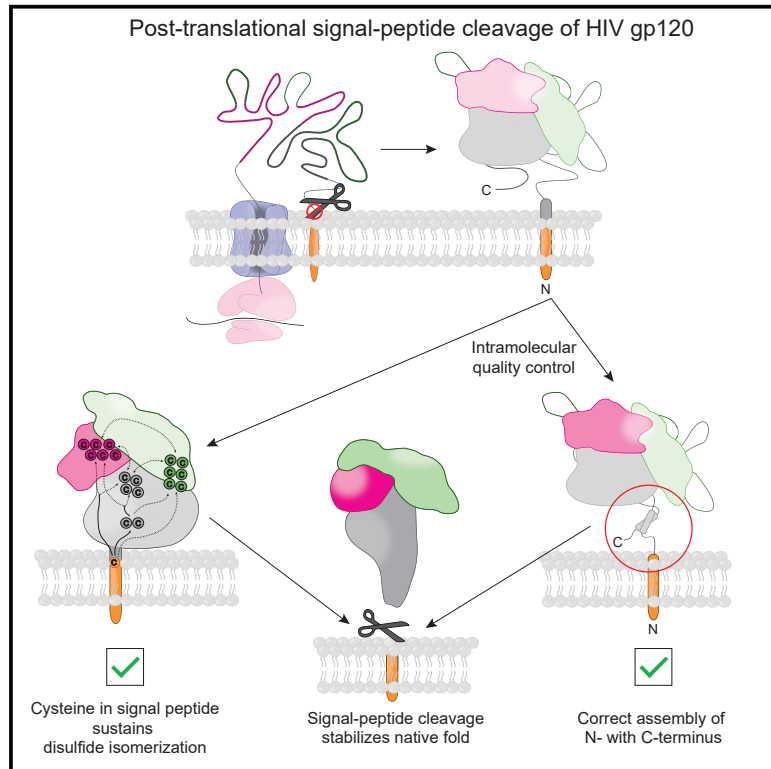


# Intramolecular quality control: HIV-1 envelope gp160 signal-peptide cleavage as a functional folding checkpoint

## Graphical abstract



## Authors

Nicholas McCaul, Matthias Quandte, Ilja Bontjer, ..., James M. Binley, Rogier W. Sanders, Ineke Braakman

## Correspondence

i.braakman@uu.nl

## In brief

McCaul et al. show that the late folding step of N and C termini integration triggers post-translational signal-peptide cleavage of HIV-1 gp120. During folding, a signal-peptide cysteine destabilizes the nearest disulfide bond, providing intramolecular disulfide isomerase activity to gp120. Signal-peptide cleavage removes the driver of isomerization, stabilizing gp120's final conformation.

## Highlights

- Intramolecular quality control via two post-targeting roles of gp120 signal peptide
- Assembly of N and C termini of HIV-1 gp120 triggers signal-peptide cleavage
- Redox-active signal-peptide cysteine sustains conformational plasticity
- Removal of signal peptide halts isomerization, stabilizing functional conformation



## Article

# Intramolecular quality control: HIV-1 envelope gp160 signal-peptide cleavage as a functional folding checkpoint

Nicholas McCaul,<sup>1,5</sup> Matthias Quandt,<sup>1</sup> Ilja Bontjer,<sup>2</sup> Guus van Zadelhoff,<sup>1</sup> Aafke Land,<sup>1</sup> Ema T. Crooks,<sup>3</sup> James M. Binley,<sup>3</sup> Rogier W. Sanders,<sup>2,4</sup> and Ineke Braakman<sup>1,6,\*</sup>

<sup>1</sup>Cellular Protein Chemistry, Bijvoet Center for Biomolecular Research, Science4Life, Faculty of Science, Utrecht University, Padualaan 8, 3584 Utrecht, the Netherlands

<sup>2</sup>Department of Medical Microbiology, Amsterdam UMC, University of Amsterdam, Amsterdam Institute for Infection and Immunity, 1105 Amsterdam, the Netherlands

<sup>3</sup>San Diego Biomedical Research Institute, 10865 Road to the Cure #100, San Diego, CA, USA

<sup>4</sup>Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, NY 10021, USA

<sup>5</sup>Present address: Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, MA, USA

<sup>6</sup>Lead contact

\*Correspondence: [i.braakman@uu.nl](mailto:i.braakman@uu.nl)

<https://doi.org/10.1016/j.celrep.2021.109646>

## SUMMARY

Removal of the membrane-tethering signal peptides that target secretory proteins to the endoplasmic reticulum is a prerequisite for proper folding. While generally thought to be removed co-translationally, we report two additional post-targeting functions for the HIV-1 gp120 signal peptide, which remains attached until gp120 folding triggers its removal. First, the signal peptide improves folding fidelity by enhancing conformational plasticity of gp120 by driving disulfide isomerization through a redox-active cysteine. Simultaneously, the signal peptide delays folding by tethering the N terminus to the membrane, until assembly with the C terminus. Second, its carefully timed cleavage represents intramolecular quality control and ensures release of (only) natively folded gp120. Postponed cleavage and the redox-active cysteine are both highly conserved and important for viral fitness. Considering the ~15% proteins with signal peptides and the frequency of N-to-C contacts in protein structures, these regulatory roles of signal peptides are bound to be more common in secretory-protein biogenesis.

## INTRODUCTION

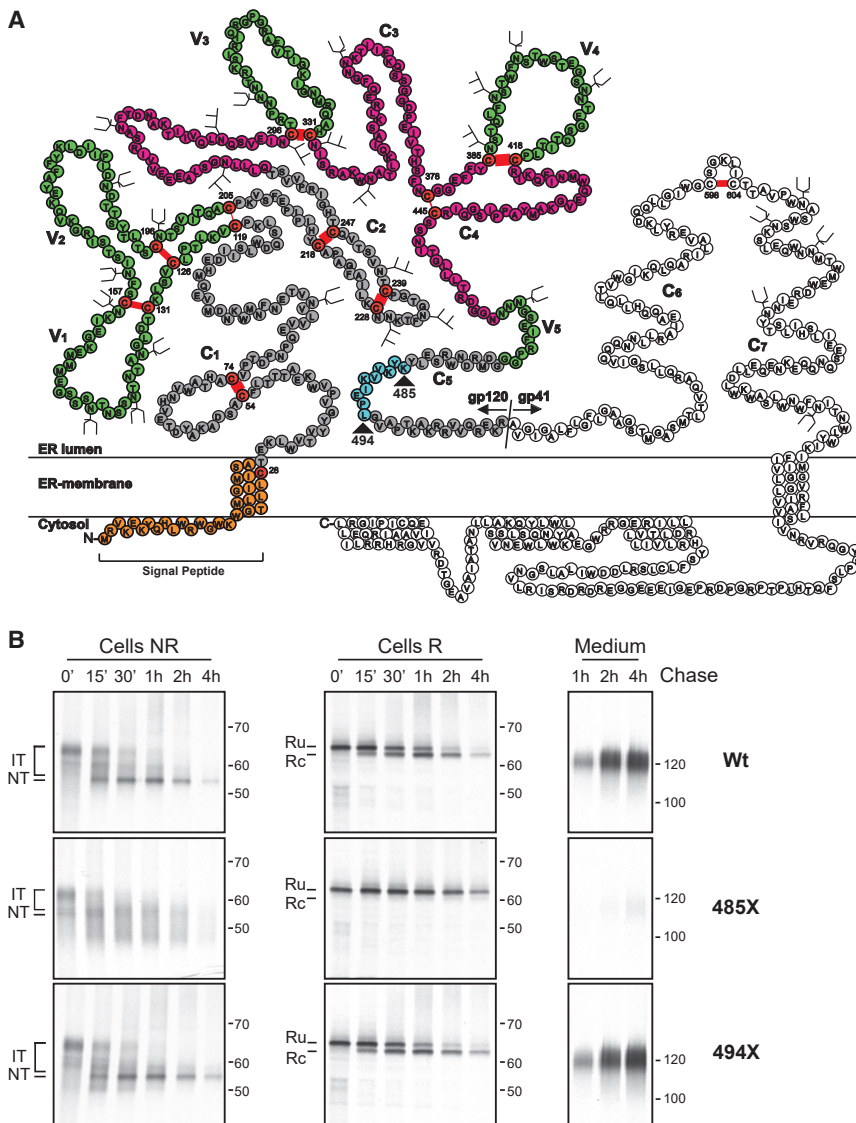
The endoplasmic reticulum (ER) is home to a wealth of resident chaperones and folding enzymes that cater to roughly a third of all mammalian proteins during their biosynthesis (Elgaard et al., 2016; Kanapin et al., 2003). It is the site of N-linked glycan addition and disulfide-bond formation, both of which contribute to protein folding, solubility, stability, and function. Targeting to the mammalian ER in general is mediated by N-terminal signal peptides, which direct the ribosome-nascent chain complex to the membrane and initiate co-translational translocation (Blobel and Dobberstein, 1975; Görlich et al., 1992a, 1992b; Jackson and Blobel, 1977; Lingappa et al., 1977; Walter and Blobel, 1981). For soluble and type-I transmembrane proteins, the N-terminal signal peptide is 15–50 amino acids long and contains a cleavage site recognized by the signal peptidase complex (von Heijne, 1985). Despite large variations in sequence, conserved features do exist. These include a positively charged N-terminal n-region, a hydrophobic h-region, and an ER-luminal c-region (von Heijne, 1983, 1984, 1985).

Paradigm-establishing studies showed that cleavable signal peptides are removed co-translationally, immediately upon

exposure of the cleavage site in the ER lumen (Blobel and Dobberstein, 1975; Jackson and Blobel, 1977). This implies that signal peptides function only as cellular postal codes and that signal-peptide cleavage and folding are independent events. However, evidence that increased nascent-chain lengths are required for cleavage is emerging (Daniels et al., 2003; Hegde and Bernstein, 2006; Rutkowski et al., 2003), indicating that the signal peptidase does not cleave each consensus site immediately upon exposure to the ER lumen. Examples are the influenza virus hemagglutinin, in which signal-peptide cleavage occurs on the longer nascent chain, well after glycosylation (Daniels et al., 2003); EDEM1 (ER degradation enhancing mannosidase-like protein) (Tamura et al., 2011); human cytomegalovirus (HCMV) proteins US11 (Rehm et al., 2001) and US9 (Seidel et al., 2021); and HIV-1 envelope glycoprotein gp160 (Li et al., 1994, 2000), suggesting that signal peptides can function as more than mere postal codes. Late signal-peptide cleavage is easily overlooked because western blots lack temporal resolution and may mask small mass differences.

Gp160 is the sole antigenic protein on the surface of the HIV-1 virion and mediates HIV-1 entry into target cells (Wyatt and





**Figure 1. Signal-peptide cleavage requires the gp120 C terminus**

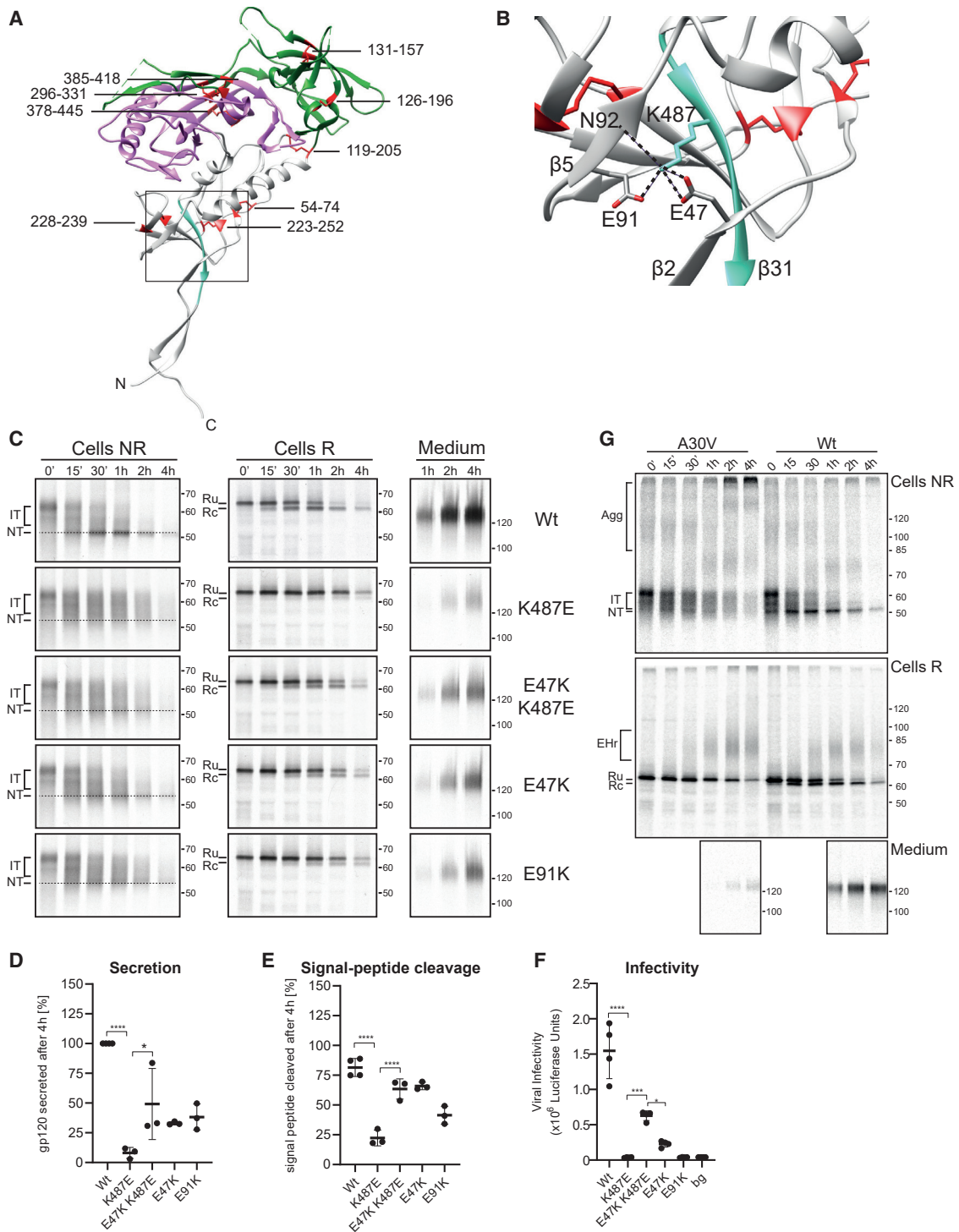
(A) Schematic representation of gp160 amino acid sequence with its signal peptide (orange) still attached (adapted from Leonard et al., 1990). Gp120 inner domain is shown in gray and the outer domain in pink (according to Pancera et al., 2010), cysteines (red) are numbered, and disulfide bonds are represented by red bars. Thickness of disulfide bonds is representative of their importance for folding and/or infectivity (thickest essential for folding, middle dispensable for folding, essential for infectivity, thinnest dispensable for both; van Anken et al., 2008). Gp120 contains five constant regions (C1–C5) and five variable regions (green, V1–V5). Oligomannose and complex glycans are represented as three- or two-pronged forked symbols, respectively. Amino acid stretch 485–494 is marked in teal.

(B) HeLa cells transiently expressing gp120 WT and C-terminal truncations were radiolabeled for 5 min and chased for the indicated times. After detergent lysis, samples were immunoprecipitated with polyclonal antibody 40336. After immunoprecipitation, samples were deglycosylated with endoH to allow for easier identification of folding ITs, analyzed by non-reducing (Cells NR) and reducing (Cells R) 7.5% SDS-PAGE. Gp120 was immunoprecipitated from medium samples with antibody 40336 and directly analyzed by reducing 7.5% SDS-PAGE (Medium). Gels were dried and exposed to Kodak-MR films or Fujifilm phosphor screens for quantification. Ru, reduced, signal peptide uncleaved gp120; Rc, reduced, signal-peptide-cleaved gp120; ITs, intermediates; NT, native.

Sodroski, 1998). It folds and trimerizes in the ER, leaves the compartment upon release by chaperones, and is cleaved by Golgi furin proteases into two non-covalently associated subunits: the soluble subunit gp120 (Figure 1A, in colors), which binds host-cell receptors, and the transmembrane subunit gp41 (Figure 1A, uncolored), which contains the fusion peptide (Decroly et al., 1994; Earl et al., 1990, 1991; Hallenberger et al., 1992; Wyatt and Sodroski, 1998). The so-called outer-domain residues (according to Pancera et al., 2010) are colored in pink (Figure 1A); the inner domain, which folds from more peripheral parts of the gp120 sequence, is in gray; and the variable loops are in green. Correct function of gp160 requires proper folding, including oxidation of the correct cysteine pairs into disulfide bonds (Bontjer et al., 2009; Land and Braakman, 2001; Land et al., 2003; Sanders et al., 2008; Snapp et al., 2017). Disulfide-bond formation and isomerization in gp160 begin co-translationally, on the ribosome-attached nascent chain, and continue long

after translation, until the correct set of ten conserved disulfide bonds has been formed (Land and Braakman, 2001; Land et al., 2003). The soluble subunit gp120 can be expressed alone and folds with highly similar kinetics as gp160 (Land et al., 2003). Signal-peptide cleavage occurs only once gp120 attains a near-native (NT) conformation and requires both N-glycosylation and disulfide-bond formation, but it is gp41 independent (Land et al., 2003; Li et al., 1996). Mutation-induced co-translational signal-peptide cleavage changes the folding pathway of gp120 and is bad for viral function (Pfeiffer et al., 2006; Snapp et al., 2017). Given the interplay between signal-peptide cleavage and gp120 folding, we set out to investigate the mechanism that drives post-translational cleavage and its relevance for gp120 folding and viral fitness.

We used kinetic oxidative-folding assays combined with functional studies on recombinant HIV strains encoding gp160 mutants and discovered additional roles for the signal peptide as quality-control checkpoint and folding mediator. A conserved cysteine in the membrane-tethered signal peptide drives disulfide isomerization in the gp120 ectodomain until gp120 folding triggers signal-peptide cleavage and release of the N terminus. We uncovered this functional, mutual regulation as an



**Figure 2. Integration of gp120 N and C termini regulates signal-peptide cleavage**

(A) Gp120 crystal structure, PDB: 5CEZ (Garces et al., 2015). Domains are colored as in Figure 1A. N and C termini are indicated; disulfide bonds are shown as red lines. Inner-domain  $\beta$  sandwich is boxed.

(B) Zoom in of inner-domain  $\beta$  sandwich. C-terminal  $\beta$  strand is shown in teal with K487 forming hydrogen bonds (dashed lines) with E47, E91, and main-chain oxygen of N92. Beta strands are numbered, and disulfide bonds are indicated as red lines. Amino acids are named and numbered according to HXB2 sequence.

(C) Experiments as in Figure 1B with HeLa cells expressing WT gp120 or indicated  $\beta$  sandwich mutants.

(legend continued on next page)

intramolecular quality control that ensures NT folding of this multidomain glycoprotein.

## RESULTS

### Signal-peptide cleavage requires the gp120 C terminus

Of the nine disulfide bonds in gp120, five are critical for proper folding and signal-peptide cleavage: three in the constant regions of gp120 in the (gray) inner domain, and two in the outer (pink) domain at the base of variable loops (green) V3 and V4 (Figure 1A; van Anken et al., 2008). Gp120 undergoes extensive disulfide isomerization during folding, as seen from the smear of gp120 folding intermediates (ITs) in the non-reducing (NR) gel upon <sup>35</sup>S-radiolabeling, from reduced gp120 down to beyond NT gp120 (Land et al., 2003; Figure 1B, NR, 0' chase). This smear disappears into a NT band with discrete mobility at around the time the signal peptide is removed (Figure 1B, R, from 15' chase). Yet, it is unclear which aspect of gp120 folding triggers signal-peptide cleavage. We therefore prepared C-terminal truncations of gp120 from 110-amino-acid length (111X) and analyzed signal-peptide cleavage (Figure S1). Radioactive pulse-chase experiments showed that only full-length gp120 (511 residues) and 494X lost their signal peptides. In all shorter forms, including 485X, the signal peptide remained uncleaved and attached to the protein (Figure S1). We continued with a time course for 485X and 494X to examine their folding pathways (Figure 1B). Both truncations contain the entire gp120 sequence except for the last 26 and 17 amino acids, respectively (Figure 1A).

Like wild-type (WT) gp120, immediately after pulse labeling (synthesis), the 2 C-terminally truncated mutants ran close to the position of reduced gp120 in non-reducing SDS-PAGE. The 485X truncation formed disulfide bonds and compacted: the folding ITs ran lower in the gel than reduced protein. It failed to form NT gp120, however (NT; Figure 1B), or another stable IT. Instead, it compacted far beyond mobility of NT and remained highly heterogeneous, suggesting formation of long-distance disulfide bonds that increased electrophoretic mobility (Figure 1B, Cells NR, 15'–2 h). Only a fraction, if any, of the 485X mutant lost its signal peptide or acquired competence to leave the ER and be secreted (Figure 1B, Medium), even though all cysteines were present in 485X gp120. The addition of only nine residues made 494X behave like WT gp120: cleavage rate and secretion were indistinguishable (Figure 1B). Oxidative folding was similar as well, except for a transient non-NT disulfide-linked pool that ran more compact than NT (Figure 1B, Cells NR, 0–1 h) and disappeared over time (Cells NR, 2–4 h). We concluded that signal-peptide cleavage required synthesis and folding of more than 484 out of the 511 amino acids of gp120. The downstream amino acids in the C5 region in the inner (gray) domain (Figure 1A, teal) triggered the switch from non-cleavable (NC) to cleavable.

### A pseudo salt bridge in the inner-domain $\beta$ sandwich controls signal-peptide cleavage and gp160 function

Amino acids 485–494 form a  $\beta$  strand (Figures 2A and 2B, teal,  $\beta$ 31) in the  $\beta$  sandwich in the inner (gray) domain of gp120 (Figures 1A, 2A, and 2B) (strand and helix coding from Garces et al., 2015). This  $\beta$  sandwich consists of seven  $\beta$  strands in constant domains C1, C2, and C5 (Figures 1A, 2A, and 2B). Six of the strands are close to the N terminus, and the seventh strand,  $\beta$ 31 (teal), is contributed by the C-terminal C5 region. As the addition of  $\beta$ 31 triggered signal-peptide removal, we hypothesized that the complete and properly folded  $\beta$  sandwich was the minimum requirement for cleavage.

To address this, we designed charge mutants aimed to prevent assembly of  $\beta$ 31 with the N-terminal part of the  $\beta$  sandwich (Figure 2B). In the gp160 structure (Garces et al., 2015), K487 (in  $\beta$ 31) forms hydrogen bonds with E47 (in  $\beta$ 2), E91 (in  $\beta$ 5), and the main-chain oxygen of N92 (in  $\beta$ 5) (Figure 2B). We made three charge-reversed mutants (E47K, E91K, and K487E) and combinations thereof (Figures 2 and S2). We excluded N92 since its interaction involves the main-chain oxygen, which cannot be removed.

As gp120 is the dominant subunit in gp160 folding and signal-peptide cleavage and allows more detailed analysis because it is smaller than gp160, we subjected WT gp120 and all mutants to pulse-chase analysis of their oxidative folding, signal-peptide cleavage, and secretion (Figures 2C–2E and S2A–S2D). As even small perturbations in folding often lead to large consequences for function (van Anken et al., 2008), we also examined the infectivity of these mutants in a luciferase-based reporter assay in TZM-bl cells (Figure 2F). Like the C-terminal truncations (485X, Figure 1B), all charge mutants in the  $\beta$  sandwich formed more compact gp120 molecules than NT gp120, implying appreciable non-NT long-range disulfide bonding, persisting at all time points or disappearing into aggregates (Figures 2C and S2B). K487E showed the strongest phenotype: minimal formation of NT and a much-delayed signal-peptide cleavage (Figures 2C, 2E and S2B, Cells NR and R, band Rc). This folding step was crucial for function, as the K487E mutant virus was non-infectious (Figure 2F).

An opposite charge reversal at the N terminus (E47K K487E) raised infectivity to ~30% of WT, significantly improving both E47K and K487E (Figure 2F). All other single, double, and triple mutants were non-infectious (Figures 2F and S2E), suggesting a rescue of infectivity of especially single-charge mutant K487E by the complementary charge reversal. Double mutants E47K K487E and E91K K487E displayed improved gp120 oxidative folding, signal-peptide cleavage, and secretion compared to K487E, but not significantly better than the E47K or E91K mutants alone (Figures 2C–2E and S2A–S2D). These N-terminal charge mutants accumulated in a NT-like band but failed to form much of the sharp NT band seen in WT (Figures 2C and S2A). Their

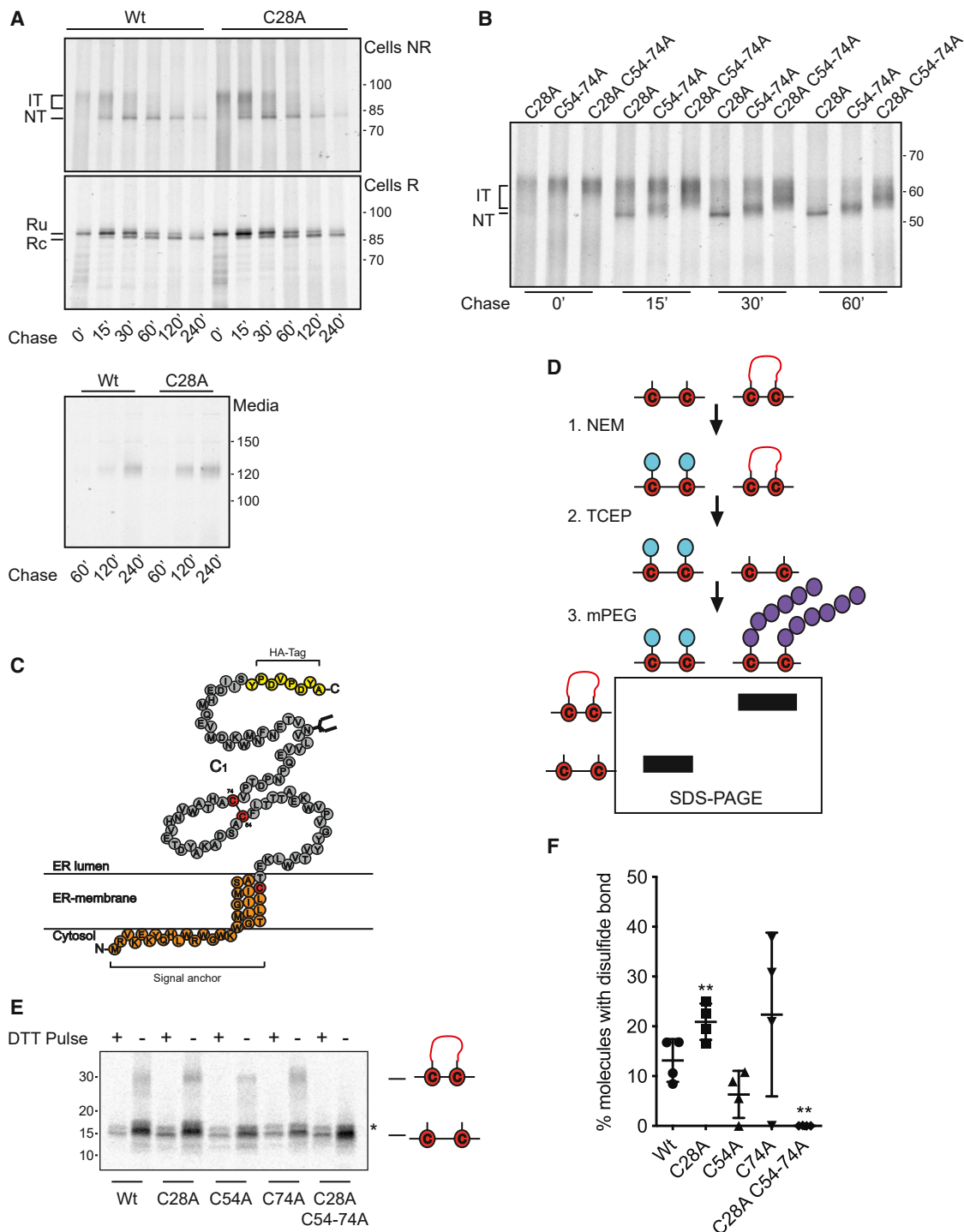
(D) Quantifications of experiments performed in (C), intracellular levels at 15' were used to correct for differences in expression between mutants and corrected values compared to WT secretion at 4 h. Error bars: SD.

(E) As in (D), but percentage of signal peptide cleaved at 4 h was measured from reducing gels.

(F) Infection assays were performed with 100 pg of HIV-1 LAI virus containing WT or mutant gp160. Error bars: SD. \**p* < 0.05, \*\*\**p* < 0.001, \*\*\*\**p* < 0.0001.

(G) Pulse chase performed as in Figure 1B. Ehr, endoH-resistant population. \*\**p* < 0.01, \*\*\*\**p* < 0.0001.

Complete statistical values are listed in Table S1.



**Figure 3. Signal-peptide cysteine is involved in gp120 oxidative folding**

(A) Pulse-chase experiments performed as in Figure 1B in HeLa cells were transfected with gp120 WT or C28A mutant.

(B) Pulse-chase experiments performed as Figure 1B in HeLa cells transfected with the indicated gp120 mutants. Cell lysates were analyzed only under non-reducing conditions.

(C) Schematic representation of gp160 111X truncation construct with its signal anchor (orange), ectodomain (gray), numbered cysteines (red), and disulfide bond indicated by red bar; C-terminal HA tag in yellow; and N-glycan depicted as forked structure.

(D) Schematic representation of mPEG alkylation-switch assay. In short, free cysteines are alkylated by NEM, which is excluded from disulfide bonds. Disulfide bonds are then reduced, and resulting free cysteines are alkylated with mPEG-maleimide, which provides a 5-kDa shift per cysteine alkylated when analyzed by SDS-PAGE.

(legend continued on next page)

signal-peptide cleavage started later and was 25%–50% lower than WT by 4 h (Cell R; [Figures 2E and S2C](#)). As a result, the secretion of all N-terminal mutants was decreased by  $\geq 60\%$  in 4 h compared to WT (Medium, [Figures 2C, 2D S2A, and S2C](#)). The redundancy of two negative charges likely contributes to the IT folding phenotype of the N-terminal mutants and the  $\sim 10\%$  residual viral infectivity of the E47K mutant ([Figure 2F](#)).

We concluded that the C-terminal  $\beta$  strand was essential for proper folding of the  $\beta$  sandwich in the inner domain, which completes folding of gp120 and triggers signal-peptide cleavage. Timing of cleavage, hence, represents a checkpoint for proper folding of gp120.

### Retention of the signal peptide causes hypercompacting of gp160

During folding, gp120 undergoes extensive disulfide formation and isomerization before reaching its NT state. These ITs appear as “waves” on SDS-PAGE representing varying degrees of compactness of folding ITs ([Land et al., 2003](#)). Because mutants of gp120 that exhibited delayed or absent cleavage all formed hypercompact forms that ran below NT ([Figures 1B and 2C](#)), we asked whether this heterogeneous electrophoretic mobility represented continued isomerization of gp120. We substituted the alanine in the  $-1$  position relative to the cleavage site for a valine to prevent cleavage by the signal-peptidase complex ([Figure 2G](#), Cell R). Initial oxidative folding of A30V was like that of WT ([Figure 2G](#), NR). However, it did not form a NT band but populated all forms, from reduced to hypercompact oxidized, probably isomerizing continuously. The monomeric forms gradually disappeared into disulfide-linked, SDS-insoluble aggregates that increased in size and eventually became too large to enter the gel ([Figure 2G](#), Agg).

In both WT and A30V gp120, a smear appeared over time at  $\sim 70$ – $80$  kDa, which we attributed to incompletely endoglycosidase H (endoH)-digested species ([Figure 2G](#), EHR). For WT gp120, this represents molecules that have transited through the Golgi complex and acquired an N-acetylglucosamine residue on their N-glycans as a first step in the modifications to complex glycan, thus acquiring endoH resistance. These complex-glycosylated gp120 molecules have yet to be secreted. In many experiments, the pool of endoH-resistant gp120 (or gp160) is too small to detect, as transport through the secretory pathway is rapid ([McCaul et al., 2019](#); [Snapp et al., 2017](#)). For overexpressed WT, but especially for A30V gp120, part of this population may be due to the inaccessibility of some glycans for removal due to formation of SDS-insoluble aggregates, visible as a smear above gp120 to the top of the gel.

We concluded that retention of the signal peptide either promotes formation of these hypercompact forms or prevents recovery from them. Because all signal-peptide-retaining mutants showed a high propensity of aggregation, it is likely that these

SDS-insoluble aggregates comprise hypercompact forms of gp120. Tethering the N terminus appears beneficial for folding, but release of gp120 from both tether and isomerization-driving cysteine is vital for stabilization of the acquired NT fold and release from the ER.

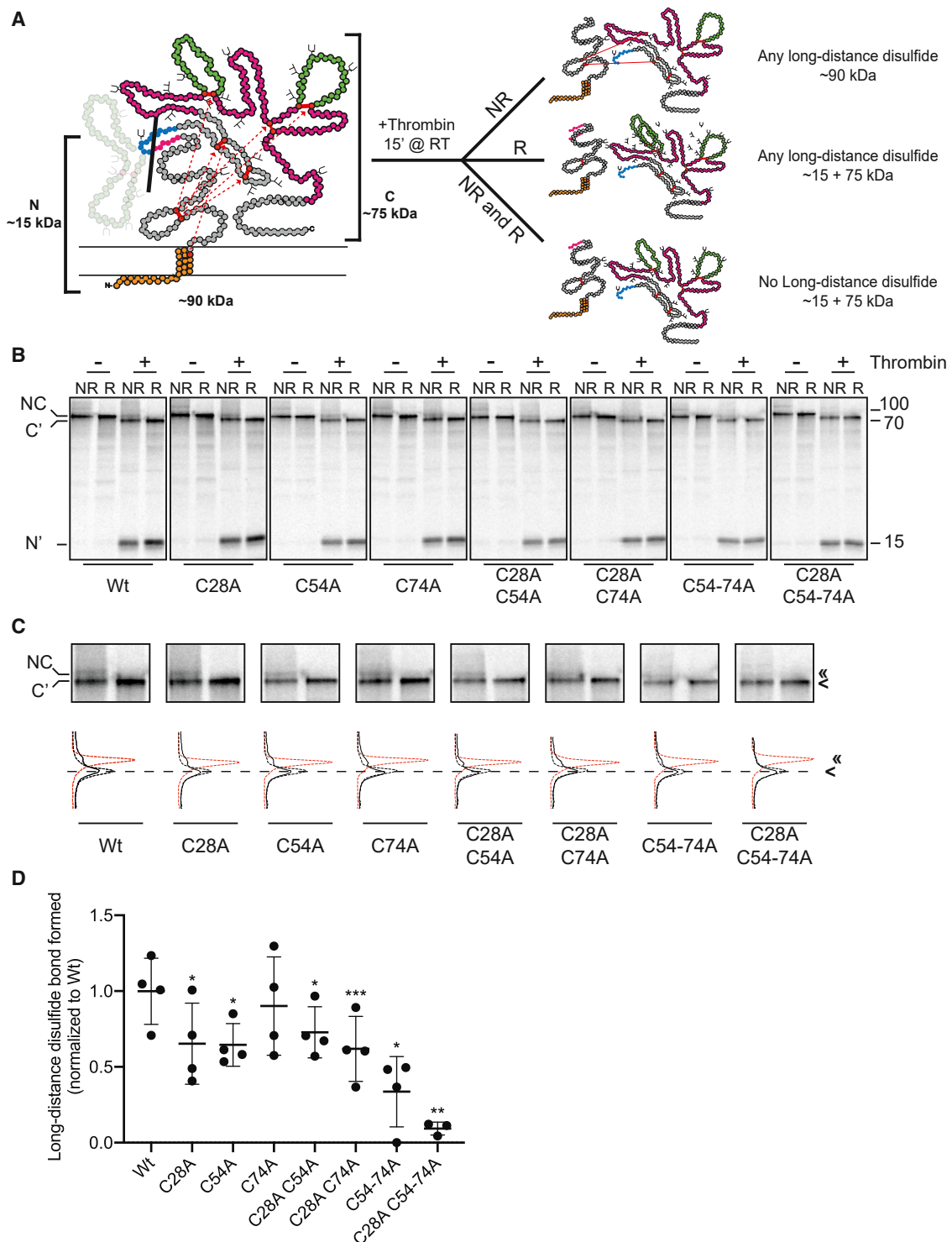
### The cysteine in the signal peptide interacts with cysteines in gp160

As the signal peptide stays attached to gp160 for at least 15 min after chain termination, it influences both co- and post-translational folding through the tethering of the N terminus to the membrane, as well as through interactions with the mature gp120 sequence ([Snapp et al., 2017](#)). The hypercompacting in NC mutants by continued disulfide isomerization ([Figure 2G](#)) implies that an unpaired cysteine must be available to keep attacking formed disulfide bonds. Opening once-formed disulfides may improve folding yield as the folding protein regains conformational freedom and, at the same time, has a chance to recover from non-NT disulfide bonding. Existing disulfides may be attacked by a cysteine from an oxidoreductase in the ER or by an intramolecular cysteine in gp120 (as shown for bovine pancreatic trypsin inhibitor [BPTI]; [Weissman and Kim, 1992, 1993](#)). The unpaired cysteine in position 28 within the signal peptide is a likely interaction candidate because it is part of the consensus sequence for the signal peptidase and, as such, (partially) exposed to the ER lumen. During translocation, C28 may interact with gp160 cysteines while they pass through the translocon. Folding analysis as in [Figures 1 and 2](#), however, showed that mutating C28 had no detectable effect on oxidative folding ([Figure 3A](#), C28A). Either C28A was identical to WT or differences are missed due to asynchrony of the folding gp120 population. To amplify mobility differences, we alkylated with iodoacetic acid, which adds a charge to each free cysteine to which it binds. To better synchronize the folding cohort, we modified the pulse-chase protocol with a preincubation with puromycin to release unlabeled nascent chains before labeling and added cycloheximide in the chase media to block elongation of radiolabeled nascent chains. The resulting radiolabeled full-length gp120 molecules all were initiated and terminated within the 5-min pulse labeling. This synchronized pool of gp120 C28A ran higher on a reducing gel ([Figure S3](#)) than synchronized WT, indicating that C28A has more free cysteines, as it bound more iodoacetic acid than WT gp120. The synchronization waned quickly during the chase, minimizing differences, but with the exception of the 2-min chase, C28A still ran higher at all time points. This persistent, reproducible difference could be due to slower disulfide-bond formation, faster disulfide-bond reduction, or a combination of both, which suggests a role for C28 in the net gp120 disulfide formation or isomerization during folding.

The importance of C28 became clear when we removed disulfide bond 54-74 in C1. Deletion of disulfide 54-74 prevents

(E) HEK293T cells expressing the indicated 111X truncations were pulse labeled for 30 min in the presence (+) or absence (–) of 5 mM DTT. At the end of the pulse, cells were scraped from dishes, homogenized, and subjected to the double-alkylation mPEG-maleimide alkylation protocol depicted in (D) ([Appenzeller-Herzog and Ellgaard, 2008](#)). After alkylation, samples were immunoprecipitated with a polyclonal antibody recognizing the HA-tag and analyzed by non-reducing 4%–15% gradient SDS-PAGE. Asterisk indicates background band.

(F) Phosphor screens from experiments performed in (E) were quantified. Error bars: SD. \*\* $p < 0.01$ , comparison to WT. Complete statistical values are listed in [Table S1](#).



**Figure 4. Gp120 exhibits long-distance, non-NT disulfide bonds during early folding**

(A) Schematic representation of gp120 thrombin-cleavable construct and the assay for long-distance disulfide bonds among the three N-terminal cysteines (C28, C54, C74) and any other cysteine in gp120. Inner domain is shown in gray, outer domain in pink, and variable loops in green, as in Figure 1A. Black bar indicates cleavage site for thrombin. Dashed red arrows indicate potential disulfide bonds.

(B) Pulse-chase experiments conducted as in Figure 1B with a 5-min pulse labeling, except that detergent lysates were immunoprecipitated with polyclonal serum HT3. After immunoprecipitation, samples were cleaved with thrombin or mock treated for 15 min at RT. All samples then were analyzed by 15%–20% discontinuous SDS-PAGE. NC: non-cleaved, full-length protein; C': C-terminal fragment; N': N-terminal fragment.

(legend continued on next page)



signal-peptide cleavage but allows gp160 to reach a compact position just above NT folded protein NT (Figure 3B; van Anken et al., 2008). When C28A was introduced into the 54-74 deletion, folding ITs were blocked at a much earlier phase and remained significantly less compact (Figure 3B). The phenotype was the same when we combined C28A with the individual deletions of C54 or C74 (Figure S4). C28A not only prevented formation of compact folding ITs, but also increased their heterogeneity. As C28 deletion aggravated folding defects of 54-74 disulfide-bond mutants, C28 must have partially compensated for the 54-74 folding defect by participating in oxidative folding. We concluded that C28A in the signal peptide was important for oxidative folding of incompletely folded gp120, most likely for sustaining isomerization of non-NT disulfide bonds, and is partially redundant with the 54-74 cysteines for this process.

To analyze whether C28, in addition to redundancy, interacted directly with the 54-74 disulfide bond, we used a 110-amino-acid truncation (111X, Figure 3C) for simplicity, as it retained the signal peptide and contains a single cysteine pair. Because formation of this disulfide bond was not detectable by comparing reduced and non-reduced samples, we made use of an alkylation-switch assay (Figure 3D; Appenzeller-Herzog and Ellgaard, 2008). In short, we radiolabeled cells expressing 111X and blocked free cysteines with *N*-ethyl maleimide (NEM). Cells then were homogenized and denatured in 2% SDS and incubated again with NEM to block any free cysteines previously shielded by structure. After immunoprecipitation and reduction of disulfide bonds with Tris 2(carboxyethyl)phosphine (TCEP), we alkylated resulting free cysteines with monomeric polyethylene glycol (mPEG)-maleimide 5,000, which adds ~5 kDa of mass for each cysteine alkylated. Samples were immunoprecipitated again to remove mPEG and were analyzed by non-reducing SDS-PAGE (Figure 3E).

The 111X construct showed only weak disulfide-bond formation, with only ~13% of molecules forming a disulfide bond (Figure 3E, WT). Upon removal of the signal-peptide cysteine C28, however, the population that contained a disulfide bond increased to ~22% (Figure 3F). The presence of C28 thus further destabilized the already-unstable 54-74 disulfide bond. The non-NT disulfide bond 28-74 in the C54A mutant barely formed, whereas the 28-54 disulfide bond in the C74A mutant was highly variable (Figure 3F). This suggests that disulfide bonds involving the signal-peptide cysteine 28 are unstable and may occur only transiently, a feature consistent with a transient role in disulfide isomerization.

### The N-terminal cysteines form long-range disulfides during early gp160 folding

Gp120 undergoes constant disulfide isomerization during folding (Land et al., 2003), and prolonged association of the

signal peptide appears to sustain isomerization and may further destabilize the already unstable 54-74 disulfide bond. Moreover, we have shown redundancy of C28 with disulfide bond 54-74, which is why we asked whether the three N-terminal cysteines in gp120 were taking part in long-range disulfide bonds during folding. We removed the V1/V2 variable loops, which are not essential for folding and function (Bontjer et al., 2009), and inserted a cleavage site for the protease thrombin (L125R). This removed all three disulfide bonds in V1/V2—126-196 and 131-157 by the loop deletion and 119-205 by mutation (C119-205A). We named the resulting construct gp120Th. Reduction after cleavage produces an N-terminal fragment of ~15 kDa containing the signal peptide and the three N-terminal cysteines C28, C54, and C74, plus an ~75-kDa fragment containing the rest of gp120 (Figure 4A). If long-range disulfides between the N- and C-terminal fragments indeed exist, the cleaved, non-reduced molecule should run in the same position as uncleaved in non-reducing conditions (~90 kDa) and should dissociate into the two fragments under reducing conditions (Figure 4A).

Pulse-chase experiments as described above were modified: instead of deglycosylation with endoH, we denatured the protein with SDS and cleaved gp120 with thrombin before separation by SDS-PAGE. As expected, gp120 that lacked all N-terminal cysteines did not form any long-distance disulfide bonds (Figures 4B and 4C, C28A C54-74A). We confirmed that WT gp120 contained long-distance disulfide bonds between N-terminal cysteines and the rest of the molecule during early folding (Figures 4B and 4C, WT). Removal of C28 significantly reduced the number of molecules with a long-range disulfide bond (Figure 4D), likely due to increased stability of the 54-74 disulfide bond in the absence of C28. Strikingly, all cysteine mutants that retained a single cysteine yielded some long-distance disulfides, suggesting that all three N-terminal cysteines could form a non-NT pair with downstream cysteines in gp120 (Figures 4B-4D).

### Removal of V1/V2 disulfides causes more rapid gp120 folding

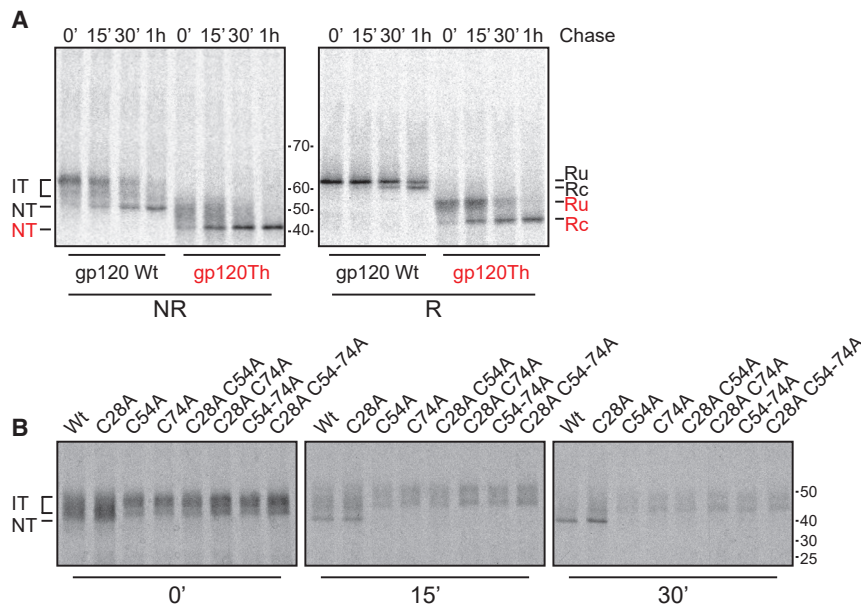
Perhaps counterintuitively, the thrombin-cleavage construct (gp120Th) folded faster than full-length gp120 (Figure 5A). Directly after the pulse, gp120Th ran as a more diffuse band, whereas full-length gp120 (gp120 WT) remained close to the reduced-gp120 mobility (Figure 5A, NR). This increased compactness shows that gp120Th had already formed more or larger-loop-forming disulfide bonds (Snapp et al., 2017). As a result, signal-peptide cleavage of gp120Th was faster—almost complete for gp120Th after a 1-h chase, compared to ~50% cleaved of gp120 WT (Figure 5A, R).

The 54-74 disulfide-bond mutants in gp120 WT background fold to a stable IT just above the NT position (Figures 3B and S4; van Anken et al., 2008), whereas the same mutants lacking

(C) Zoom in of gels from (B) showing full-length and C-terminal fragments. Lane profiles were generated from autoradiographs in ImageQuant TL. Black solid lines are from non-reduced, thrombin-cleaved samples; black dashed lines are from reduced, thrombin-cleaved samples; and red dashed lines are from non-reduced, NC samples. <, C' fragment; <<, NC fragment.

(D) Quantifications of phosphor screens from (B). Values were calculated by dividing the signal in the N-terminal fragment by the full-length uncleaved protein and subtracting the value for reducing conditions to determine percent of molecules with a long-distance disulfide bond. Resulting values then were normalized to WT and statistically compared to WT. Error bars: SD. \**p* < 0.05, \*\**p* < 0.01, \*\*\**p* < 0.001.

Complete statistical values are listed in Table S1.



**Figure 5. Folding of thrombin-cleavable gp120 construct**

(A) Pulse-chase experiments conducted as in Figure 1B, except that HeLa cells were transfected with WT full-length gp120 (gp120 WT) or thrombin-cleavable gp120 (gp120Th) and pulsed for 10 min. (B) Pulse-chase experiments conducted as in Figure 1B, except that HeLa cells were transfected with cysteine mutants of gp120Th. Polyclonal serum HT3 was used for immunoprecipitation from detergent lysates, and samples were analyzed by 7.5% non-reducing SDS-PAGE. Red text in (A) refers to gp120Th running positions.

V1/V2 (in gp120Th) failed to accumulate in a single band (Figure 5B), reminiscent of the folding of gp120 C28A C54-74A (Figure 3B). This indicates that V1/V2 deletion phenocopied C28 deletion in the 54-74 mutants. Redundancy of V1/V2 with C28 was confirmed by the lack of additional effect of C28 removal in the gp120Th 54-74 mutants (Figure 5B).

### C28A results in decreased HIV-1 production and pseudovirus infectivity

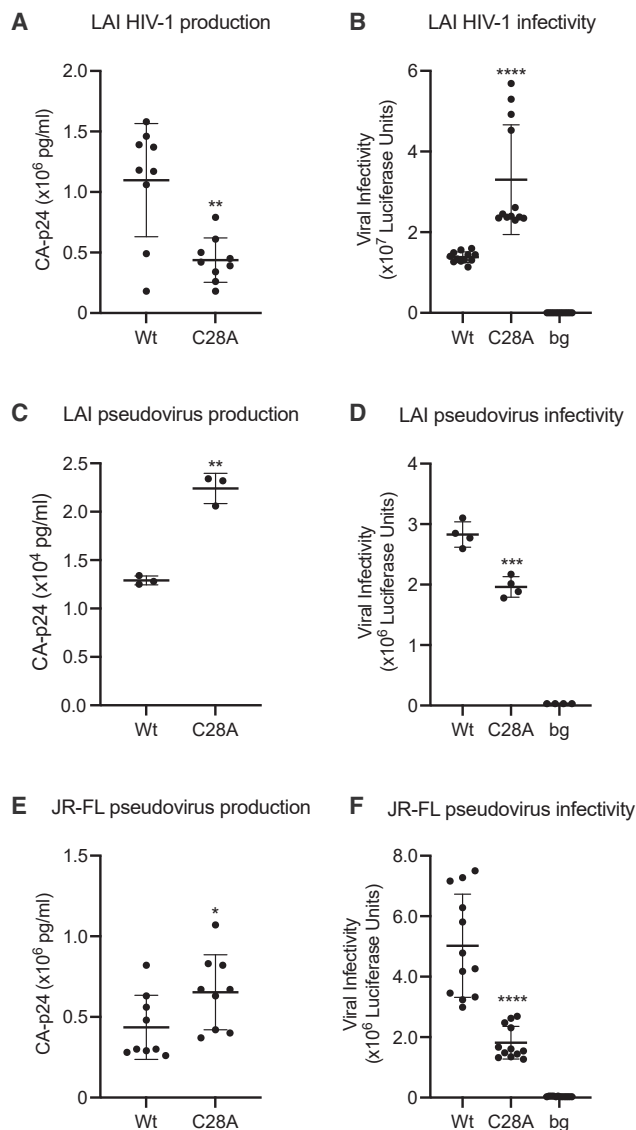
As the biochemical data suggested a role for C28 in gp120 folding, we examined the effect of C28A mutation on viral production and infectivity. For this, we transfected cells with a molecular clone containing the full HIV-1 genome (LAI strain), containing either WT gp160 or the C28A mutant. As the reading frames of HIV-1 *Env* and *Vpu* overlap, mutations in the signal peptide of gp160 can cause changes in the C terminus of Vpu, so we produced the viruses in HEK293T cells, which are deficient in CD4 and tetherin and therefore do not require Vpu to enhance virus production (Van Damme et al., 2008). C28A HIV virus was consistently produced at lower quantities than WT HIV, as measured by levels of the viral capsid protein p24 (Figure 6A). We then infected TZM-bl cells with identical quantities of the produced viruses (by CA-p24 ELISA) in the same infectivity assay as above. Strikingly, the C28A virus was significantly more infectious than WT HIV but displayed strong heterogeneity in infectivity, suggestive of heterogeneity in C28A gp160 incorporation into the virions (Figures 6B, S5C, and S5D). Because of the deficit in virus production, despite increased infectivity, C28A-gp160-containing HIV is probably less competitive in nature. Indeed, alignments of >4,300 gp160 sequences from across all subtypes show that C28 is ~87% conserved (<https://www.hiv.lanl.gov>).

To uncouple differences in virus production from infectivity, we moved to the pSG3<sup>ΔEnv</sup> pseudovirus system. pSG3<sup>ΔEnv</sup> is an HIV-1 clone with a defective *Env* gene that requires supple-

mentation with exogenous Env to produce infectious virions (Wei et al., 2002). This allows analysis of the effect of C28A gp160 on infectivity alone. In the pseudovirus system, production of C28A virus was higher than that of the WT virus (Figure 6C). Infectivity of LAI-C28A-gp160-bearing pseudovirus, however, was reduced by ~25% compared to WT (Figures 6D, S5D, and S5E). The LAI strain of HIV shows near-complete shedding of gp120 from gp41 (Earl et al., 1991; Land and Braakman, 2001; Land et al., 2003). Many of the residues responsible for the gp120-41 interaction are found in the inner-domain β sandwich that we identified as crucial for regulating both signal-peptide cleavage and infectivity. Because of this, we considered whether these results would hold in a less labile isolate. We therefore assessed the JR-FL isolate, which has a stronger gp120-41 interaction and barely sheds gp120. We found very little difference in pseudovirus production between WT and C28A JR-FL gp160 (Figure 6E). Infectivity of the C28A gp160 JR-FL pseudovirus was roughly 60% less than the infectivity of WT (Figures 6F, S5G, and S5H). Finally, to control for differing amounts of gp160 incorporation into viruses, we produced virus-like particles containing various signal-peptide mutants of gp160 in the JR-FL strain and measured their gp160 content by BN-PAGE followed by western blot (Figure S6). Whereas the co-translational mutant M26P (Snapp et al., 2017) resulted in decreased gp160 incorporation, C28A gp160 was incorporated at similar levels to WT (Figure S6). We concluded, therefore, that gp160 conformation, its function, and, as a result, HIV suffered from the removal of the signal-peptide cysteine.

### DISCUSSION

During synthesis and for 15–30 min afterward, the N terminus of HIV-1 gp160 remains tethered to the ER membrane by its transient signal anchor. We show that conformational plasticity is enhanced through the cysteine in the signal peptide driving disulfide isomerization, in part via the 54-74 disulfide bond, until the gp120 C and N termini have assembled, completing the inner-domain β sandwich and gp120 folding (Figure 7). This triggers signal-peptide cleavage, removing C28 from the protein, halting further isomerization, and stabilizing NT gp120. This



**Figure 6. C28A gp160 is detrimental to HIV-1 production and pseudovirus infectivity**

(A) HEK293T cells were transfected with WT or C28A mutant pLAI constructs, and virus production was measured by CA-p24 ELISA. Error bars: SD, \*\* $p < 0.01$ .

(B) Infection assays were performed as in Figure 2F with LAI HIV-1 virus containing WT or mutant gp160, as produced in (A). Error bars: SD, \*\*\*\* $p < 0.0001$ .

(C) Virus produced as in (A), except C33A cells were co-transfected with WT or C28A mutant LAI gp160 constructs in a 2:1 ratio with pSG3 $\Delta$ Env construct (pseudovirus). Error bars: SD, \*\* $p < 0.01$ .

(D) Infectivity assays were performed as in Figure 2F. Error bars: SD, \*\*\* $p < 0.001$ .

(E) Virus produced as in (A), except cells were co-transfected with WT or C28A mutant JR-FL gp160 constructs in a 2:1 ratio with pSG3 $\Delta$ Env construct (pseudovirus). Error bars: SD.

(F) Infectivity assays were performed as in Figure 2F. Error bars: SD, pSG3 $\Delta$ Env: virus produced without JR-FL gp160 construct, \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\*\* $p < 0.0001$ .

Complete statistical values are listed in Table S1.

intramolecular quality-control process is essential for HIV fitness and can be impaired and restored by single charge reversals in the gp120 inner domain.

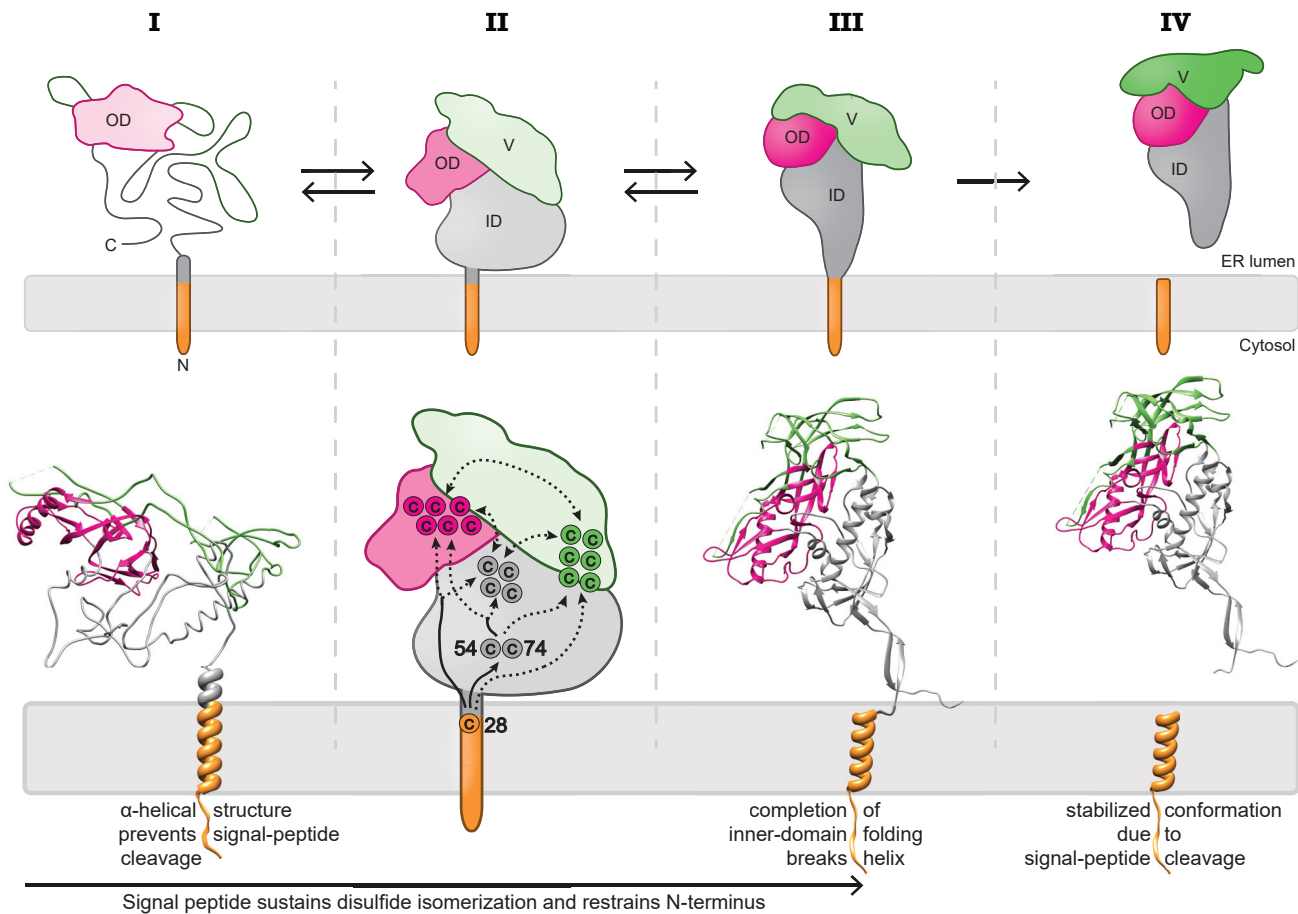
### Hierarchy of gp120 folding

The inner domain with the  $\beta$  sandwich and the outer domain, which contains a stacked double- $\beta$  barrel, together constitute the minimal folding-capable “core” of gp120 (Figure 2A; Garces et al., 2015; Kwong et al., 1998). Gp120 is completed with the surrounding variable loops V1/V2, V3, and V4 (Figures 1A, 2A, and 2B). The core contains six of the nine disulfide bonds in gp120, including the five that are essential for correct folding and signal-peptide cleavage (van Anken et al., 2008). The inner-domain  $\beta$  sandwich consists of seven strands, six of which are N termini. Proper folding of the sandwich requires assembly with the C-terminal  $\beta$ 31 strand and formation of the five essential disulfide bonds, which then leads to signal-peptide cleavage (Figure 2A; van Anken et al., 2008).

Some folding of the gp120 “hairpin” begins during translation and translocation into the ER (Land et al., 2003), but the bulk occurs post-translationally (Figure 7). The outer domain is the first complete domain to emerge from the translocon and has low contact order, meaning that folding does not require integration of distal residues (Figure 7). It is the first domain to fold, which requires formation of the two NT disulfide bonds (296-331 and 385-418) in the  $\beta$  barrel: gp120 lacking either disulfide bond barely folds past the reduced position in SDS-PAGE (Sanders et al., 2008; van Anken et al., 2008).

The inner domain of gp120 folds next: individual deletions of its three essential disulfide bonds (54-74, 218-247, and 228-239) fold into more compact structures than the outer-domain deletions (van Anken et al., 2008). The most intriguing is the 54-74 disulfide: the C54-74A mutant accumulates in a sharp band just above the NT position, retaining its signal peptide. In contrast, C218-247A and C228-239A fail to form defined ITs (van Anken et al., 2008), suggesting that these  $\beta$ -sandwich-embracing disulfides (Figure 2A) stabilize the inner domain.

Folding of the inner domain leads to cleavage of the signal peptide. Until that time, the signal peptide acts as signal anchor because it adopts an  $\alpha$ -helical conformation that extends past the cleavage site and prevents proteolytic cleavage (Snapp et al., 2017). Folding and integration of the inner domain must break the helix to allow cleavage, as in crystal structures of gp120, this early helical region is a  $\beta$  strand (Garces et al., 2015). As the disulfide bonds of V1/V2 and gp41 are dispensable for signal-peptide cleavage and their deletion shows no aberration in the folding pathway (van Anken et al., 2008), those domains likely fold after or independent of the outer and inner domains (Figure 7). This is underscored by gp120 folding and signal-peptide cleavage being largely independent of gp41 (Land et al., 2003) and by N- and C-terminal sequences in the gp120 inner domain forming the binding site of gp41 (Garces et al., 2015; Julien et al., 2013; Lyumkis et al., 2013). Gp41 binding may explain apparent inconsistencies between folding and function of some inner-domain mutants (Garces et al., 2015; Yang et al., 2003). The conserved gp41 binding site on the gp120 inner-domain  $\beta$  barrel also may explain the conservation (and hence value) of the intramolecular quality-control



**Figure 7. Model for post-translational gp120 folding, signal-peptide cleavage, and intramolecular disulfide isomerization**

Shown are the conformational changes in the signal peptide and proximal areas during gp120 folding that lead to cleavage. (I) At the end of translation, gp120 is far from folded;  $\alpha$ -helical structure around the signal peptide prevents signal-peptide cleavage. (II) As long as the signal-peptide remains attached, C28 sustains intramolecular disulfide isomerization by interacting with downstream cysteine residues. N-terminal tethering limits conformational freedom, which may support NT folding. Solid lines, interactions found experimentally; dashed lines, predicted interactions. (III) Folding and integration of the inner-domain  $\beta$  sandwich breaks the signal-peptide helix, allowing access to the consensus peptide cleavage site. (IV) Signal-peptide cleavage stabilizes the conformation of gp120 by removal of the free sulfhydryl group of C28, thus preventing continued disulfide isomerization. Gray, inner domain; pink, outer domain; green, variable loops; orange, signal peptide.

system: it ensures proper folding of this binding site, well timed and with high fidelity, before gp41 folding.

The regulation of signal-peptide cleavage by folding of gp120 implies that  $\beta$ -sandwich formation generates sufficient force to break the  $\alpha$ -helix and expose the cleavage site.  $\alpha$ -helical proteins have lower mechanical strength than  $\beta$ -sheet proteins, which often need to resist dissociation and unfolding; lower mechanical strength facilitates conformational changes to expose transient binding sites or to allow signaling (Chen et al., 2015). Forces of  $\sim 5$ – $8$  pN have been reported as sufficient to break an  $\alpha$ -helix (del Rio et al., 2009; Gordon et al., 2015; Yao et al., 2014; Zhang et al., 2009), while pulling apart a  $\beta$ -sheet protein such as immunoglobulin (Ig) domains, outer surface protein A (OspA), or ubiquitin by shearing needs  $>160$  pN (Brockwell et al., 2003; Carrion-Vazquez et al., 2003; Hertadi et al., 2003), although less force is required for peeling (Brockwell et al., 2003). Physiological forces measured so far do not exceed  $\sim 40$  pN, a force

level at which proteins in general may be destabilized already (Chen et al., 2015). The  $\alpha$ -helical region around the cleavage site in gp120, thus, would lose the stability competition from the  $\beta$ -sheet formation in the inner domain if their structures are incompatible (Figures 2A and 7). First-time folding (i.e., the completion of the inner-domain  $\beta$  sandwich by assembly of  $\beta 31$ ) is likely to generate sufficient force as well, as 7–12 pN allows constant binding of a filamin  $\beta$  strand to a  $\beta$  sheet (Rognoni et al., 2012). Formation of the inner-domain disulfide bonds may further raise the stabilizing force (Eckels et al., 2019). The completion of gp120 folding, therefore, likely generates the  $\sim 5$ -pN force needed to break the  $\alpha$ -helix and allow the signal peptidase to cleave off the gp120 signal peptide.

#### Effects of the attached signal peptide

The postponed cleavage of the signal peptide makes it a transient signal anchor, which acts as a membrane tether. This limits

conformational freedom and benefits folding: the integration of the C-terminal  $\beta$ 31 strand into the folded inner-domain  $\beta$  sandwich, as in knotted proteins (Soler and Faisca, 2012).

The prolonged proximity of the free signal-peptide cysteine (28) supports disulfide isomerization and increases conformational plasticity during gp120 folding. Gp120 requires a NT set of disulfide bonds to attain its functional 3D structure, but already during synthesis non-NT disulfides are formed, which reshuffle over time (Land et al., 2003). Despite extensive isomerization during folding, WT gp120 only transiently occupied forms more compact than NT. In contrast, the various  $\beta$  sandwich and un-cleavable mutants extensively populated hypercompact states with non-NT long-range disulfide bonds, indicating that without stable assembly of the N and C termini and with the resulting retention of the signal peptide, isomerization continues unabated, driving the formation of these hypercompact structures.

The constant disulfide isomerization is sustained by the redox-active cysteine 28 in the signal peptide, as its sulfhydryl group is free to attack existing disulfide bonds. Once gp120 folding has reached a state where isomerization is no longer preferred (N and C termini in the inner domain assembled), the signal peptide is cleaved, removing an important, conserved driving force behind isomerization. Cleavage of the signal peptide acts as a sink because it removes the disulfide-attacking cysteine and pulls the folding equilibrium to the NT structure.

#### Mode of action of C28

In a short construct, C28 favored a disulfide bond with C54 (Figure 3F). Despite a limited ability to form disulfide bonds with cysteines downstream of the 54-74 bond (Figure 4D), deletion of C28 in the absence of V1/V2 did not aggravate the C54-74A-defective phenotype. C28, hence, likely sustains isomerization in this construct primarily by constant attack and destabilization of the 54-74 disulfide bond. This propagates a free sulfhydryl group downstream through the folding protein, as the result of an intramolecular electron-transport chain from the more C-terminal cysteines via C54 and C74 to C28 (Figure 7). In the 110-residue gp120 chain (111X), essentially a mimic of a released gp120 nascent chain, the 54-74 bond had formed in maximally 20% of molecules, demonstrating its inherent instability as well as the likelihood that C28 already acts on 54 and 74 during translation.

In the presence of V1/V2, C28 showed redundancy with the 54-74 disulfide, implying that C28 can fulfill roles otherwise played by C54 and C74 (and vice versa). This suggests that C28 is involved in folding (and isomerization of V1/V2) and may play this role by direct interaction with V1/V2 cysteines, in the absence of C54 and C74, distinct from its more 54-74-mediated role in downstream disulfide bonds' formation. We cannot exclude that the attack of C28 on the V1/V2 disulfides leads to an alternative electron transport chain from C-terminal cysteines via V1/V2 to C28. Either way, the redox-active C28 needs to be removed at the end of gp120 folding to ensure stability of the gp120 conformation.

#### Intramolecular oxidoreductase and quality control for proper folding

Built-in isomerase activity may seem redundant considering that gp120 folds in the ER, a compartment that contains >20 protein-disulfide-isomerase family members (Jansen et al., 2012). Extensive

gp120 isomerization (Land et al., 2003) is undoubtedly supported by these oxidoreductases. ERp57, for instance, may well be recruited by the conserved N-linked glycosylation site N88, in close proximity to the 54-74 disulfide bond, via its interaction with lectin chaperones calnexin and calreticulin (Jansen et al., 2012; Oliver et al., 1997). While the N88 glycan plays a role in shielding the fusion peptide in gp41 from neutralizing antibodies (van Gils et al., 2016), its removal has shown only a modest effect on infectivity (Lee et al., 1992). Rather, bypass of the entire calnexin/calreticulin chaperone machinery via  $\alpha$ -glucosidase-I inhibition still allows production of infectious virions (Fenouillet et al., 1997). Despite decreasing internal oxidoreductase activity, removal of C28 showed no detectable effect on glycosylation or glycan modification, as seen by the mobility of secreted gp120 (Figure 3A).

Despite their likely role in gp160 folding, oxidoreductases are large, bulky proteins that cannot catalyze disulfide-bond formation once areas in gp160 have attained significant tertiary structure. Isomerase activity built into the folding protein allows free-thiol propagation inside the protein, even involving cysteines unreachable by folding enzymes. This, perhaps, should not be too surprising given that in folded proteins, the majority of cysteines are solvent inaccessible (Srinivasan et al., 1990), and during folding, disulfide bonds become resistant to reduction with DTT (Tatu et al., 1993, 1995).

An example of intramolecular disulfide isomerization is the cysteine in the pro-peptide of BPTI, which increased both the rate and yield of BPTI folding (Weissman and Kim, 1992). The majority of disulfide formation during *in vitro* folding of BPTI results from intramolecular disulfide rearrangements (Creighton et al., 1993; Darby et al., 1995; Weissman and Kim, 1995). Transfer of free thiols between luminal and transmembrane domains in the ER has been demonstrated for vitamin-K-epoxide reductase (Liu et al., 2014; Schulman et al., 2010), indicating that such exchanges are possible. While C28 is located in the transmembrane  $\alpha$ -helix (Snapp et al., 2017), suggesting immersion in the membrane, sliding of transmembrane domains up and down in the membrane is possible (Borochoy and Shinitzky, 1976; Danielson et al., 1994; Mowbray and Koshland, 1987). As C28 is part of the consensus sequence for signal-peptide cleavage, it likely is exposed to the ER lumen at least part of the time.

Intramolecular oxidoreductase activity offers not only an advantage but also intramolecular quality control. Release of a protein from the ER requires its folding to the extent that chaperones do not bind any more. The intramolecular quality control we here describe ensures a much more subtle regulation of conformational quality. Gp160's function as HIV-1 fusion protein requires the NT interaction between gp41 and the gp120 inner domain. Only proper exposure of the gp41 binding site in gp120 will lead to a functional protein. Intramolecular quality control ensures precision to the level of single residues as well as precision of timing, increasing robustness of the folding process.

#### Conserved and multiple roles for signal peptides

Post-translational signal-peptide cleavage of gp120 is conserved across different subtypes of HIV-1 as biochemical properties, even if sequences are not strictly conserved (Snapp et al., 2017). This appears to be more general, as in other organisms, signal peptides mutate at a lower rate (Morrison et al., 2003;

Williams et al., 2000) or with an increased proportion of null (Veitia and Caburet, 2009) or function-preserving mutants (Garcia-Maroto et al., 1991), compared to the surrounding mature peptide. Function-altering mutants often have deleterious effects (Bonfanti et al., 2009; Piersma et al., 2006).

Detailed kinetic analysis of signal-peptide cleavage has not been reported for a great number of proteins, but gp160 is not alone in its biosynthesis-dependent and biosynthesis-regulated signal-peptide cleavage (Anjos et al., 2002; Daniels et al., 2003; Matczuk et al., 2013; Rutkowski et al., 2003; Zschenker et al., 2001). Whereas structure regulates cleavage in HCMV US11 (Rehm et al., 2001; Tamura et al., 2011), function regulates cleavage in EDEM1 (Rehm et al., 2001; Tamura et al., 2011). These studies demonstrate that a variety of conditions including nascent-chain length and N-glycan addition can play a role in signal-peptide cleavage. Signal peptides are more than address labels, and signal-peptide cleavage and folding are more interdependent than originally thought.

We argue that late signal-peptide cleavage may be much more common than biochemical experiments have uncovered. Cleavage may occur at any time from co-translationally until late post-translationally. Considering the low rate of protein synthesis, ~3 to 6 amino acids per second (Braakman et al., 1991; Horwitz et al., 1969; Ingolia et al., 2011; Knopf and Lamfrom, 1965), it can lead to long average synthesis times (~1.5–3 min for gp120 and ~2–5 min for gp160). In fact, translation rates are much more heterogeneous (Ingolia et al., 2019): nascent chains of influenza virus HA may take >15 min to complete, corresponding to a rate of less than one residue per second (Braakman et al., 1991; unpublished observations). For large proteins, the difference between early and late co-translational cleavage leaves a window of several minutes, during which the signal peptide functions as an anchor tethering the protein to the ER membrane. Sequence features in the signal peptide, such as an exposed cysteine, are given the opportunity to interact with the folding protein.

The membrane tether limits conformational freedom of the protein and reduces overall conformational entropy, which is predicted to increase fidelity of protein folding and stability (Dill and Alonso, 1988; Zhou, 2008; Zhou and Dill, 2001). This may well benefit the formation of N- and C-terminal contacts in proteins, which are present in ~50% of soluble PDB structures (Krishna and Englander, 2005) and are present in multiple multimeric viral glycoproteins (Chen et al., 1998; Garces et al., 2015; Gogala et al., 2014; Sauter et al., 1992; Sun et al., 2014).

Here, we have presented compelling evidence for the direct functional contribution of the signal peptide to HIV-1 gp160 folding. The signal peptide drives disulfide isomerization of gp120 during folding, increasing conformational plasticity while tethering the N terminus, and functions as quality control organizer, leaving only after near-NT conformation has been attained. As evidence grows, it becomes clear that signal peptides demonstrate functions far beyond their originally assigned roles as cellular postal codes.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Cell lines
  - Viruses
- **METHOD DETAILS**
  - Plasmids, antibodies and reagents
  - Transient transfection
  - Virus production
  - Infectivity assay
  - Folding assay
  - Deglycosylation, SDS-PAGE, and autoradiography
  - mPEG treatment
  - Thrombin cleavage
  - Native PAGE and western blotting
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2021.109646>.

## ACKNOWLEDGMENTS

We would like to thank members of the Braakman-Van der Sluijs and Sanders labs for their fruitful discussions and insights—in particular, Peter van der Sluijs for critical reading of the manuscript and Joseline Houwman for critical reading of the manuscript and design of Figure 7. This work was supported by grants from the Dutch Research Council (NWO)- Chemical Sciences 711.012.008 (I. Braakman, N.M., A.L., M.Q.), the European Union 7th framework program, ITN “Virus Entry” 235649 (I. Braakman, N.M., M.Q.), and the European Union’s Horizon 2020 research and innovation program under grant 681137 (R.W.S. and I. Bontjer). R.W.S. is a recipient of a Vici grant from the Dutch Research Council (NWO: 91818627).

## AUTHOR CONTRIBUTIONS

Conceptualization, N.M., M.Q., and I. Braakman; methodology, N.M.; investigation, N.M., M.Q., I. Bontjer, and A.L.; writing – original draft, N.M., M.Q., and I. Braakman; writing – review & editing, N.M., M.Q., I. Bontjer, R.W.S., A.L., and I. Braakman; funding acquisition, R.W.S. and I. Braakman.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 12, 2021

Revised: June 28, 2021

Accepted: August 11, 2021

Published: August 31, 2021

## REFERENCES

- Anjos, S., Nguyen, A., Ounissi-Benkhalha, H., Tessier, M.C., and Polychronakos, C. (2002). A common autoimmunity predisposing signal peptide variant of the cytotoxic T-lymphocyte antigen 4 results in inefficient glycosylation of the susceptibility allele. *J. Biol. Chem.* 277, 46478–46486.
- Appenzeller-Herzog, C., and Ellgaard, L. (2008). In vivo reduction-oxidation state of protein disulfide isomerase: the two active sites independently occur in the reduced and oxidized forms. *Antioxid. Redox Signal.* 10, 55–64.

- Blobel, G., and Dobberstein, B. (1975). Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *J. Cell Biol.* **67**, 835–851.
- Bonfanti, R., Colombo, C., Nocerino, V., Massa, O., Lampasona, V., Iafusco, D., Viscardi, M., Chiumello, G., Meschi, F., and Barbetti, F. (2009). Insulin gene mutations as cause of diabetes in children negative for five type 1 diabetes autoantibodies. *Diabetes Care* **32**, 123–125.
- Bontjer, I., Land, A., Eggink, D., Verkade, E., Tuin, K., Baldwin, C., Pollakis, G., Paxton, W.A., Braakman, I., Berkhout, B., and Sanders, R.W. (2009). Optimization of human immunodeficiency virus type 1 envelope glycoproteins with V1/V2 deleted, using virus evolution. *J. Virol.* **83**, 368–383.
- Bontjer, I., Melchers, M., Eggink, D., David, K., Moore, J.P., Berkhout, B., and Sanders, R.W. (2010). Stabilized HIV-1 envelope glycoprotein trimers lacking the V1V2 domain, obtained by virus evolution. *J. Biol. Chem.* **285**, 36456–36470.
- Borochoy, H., and Shinitzky, M. (1976). Vertical displacement of membrane proteins mediated by changes in microviscosity. *Proc. Natl. Acad. Sci. USA* **73**, 4526–4530.
- Braakman, I., Hoover-Litty, H., Wagner, K.R., and Helenius, A. (1991). Folding of influenza hemagglutinin in the endoplasmic reticulum. *J. Cell Biol.* **114**, 401–411.
- Brockwell, D.J., Paci, E., Zinober, R.C., Beddard, G.S., Olmsted, P.D., Smith, D.A., Perham, R.N., and Radford, S.E. (2003). Pulling geometry defines the mechanical resistance of a beta-sheet protein. *Nat. Struct. Biol.* **10**, 731–737.
- Carrion-Vazquez, M., Li, H., Lu, H., Marszalek, P.E., Oberhauser, A.F., and Fernandez, J.M. (2003). The mechanical stability of ubiquitin is linkage dependent. *Nat. Struct. Biol.* **10**, 738–743.
- Cham, F., Zhang, P.F., Heyndrickx, L., Bouma, P., Zhong, P., Katinger, H., Robinson, J., van der Groen, G., and Quinnan, G.V., Jr. (2006). Neutralization and infectivity characteristics of envelope glycoproteins from human immunodeficiency virus type 1 infected donors whose sera exhibit broadly cross-reactive neutralizing activity. *Virology* **347**, 36–51.
- Chen, J., Lee, K.H., Steinhauer, D.A., Stevens, D.J., Skehel, J.J., and Wiley, D.C. (1998). Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation. *Cell* **95**, 409–417.
- Chen, Y., Radford, S.E., and Brockwell, D.J. (2015). Force-induced remodeling of proteins and their complexes. *Curr. Opin. Struct. Biol.* **30**, 89–99.
- Creighton, T.E., Bagley, C.J., Cooper, L., Darby, N.J., Freedman, R.B., Kemmink, J., and Sheikh, A. (1993). On the biosynthesis of bovine pancreatic trypsin inhibitor (BPTI). Structure, processing, folding and disulphide bond formation of the precursor in vitro and in microsomes. *J. Mol. Biol.* **232**, 1176–1196.
- Crooks, E.T., Tong, T., Osawa, K., and Binley, J.M. (2011). Enzyme digests eliminate nonfunctional Env from HIV-1 particle surfaces, leaving native Env trimers intact and viral infectivity unaffected. *J. Virol.* **85**, 5825–5839.
- Daniels, R., Kurovski, B., Johnson, A.E., and Hebert, D.N. (2003). N-linked glycans direct the cotranslational folding pathway of influenza hemagglutinin. *Mol. Cell* **11**, 79–90.
- Danielson, M.A., Biemann, H.P., Koshland, D.E., Jr., and Falke, J.J. (1994). Attractant- and disulfide-induced conformational changes in the ligand binding domain of the chemotaxis aspartate receptor: a 19F NMR study. *Biochemistry* **33**, 6100–6109.
- Darby, N.J., Morin, P.E., Talbo, G., and Creighton, T.E. (1995). Refolding of bovine pancreatic trypsin inhibitor via non-native disulphide intermediates. *J. Mol. Biol.* **249**, 463–477.
- Das, A.T., Land, A., Braakman, I., Klaver, B., and Berkhout, B. (1999). HIV-1 evolves into a nonsyncytium-inducing virus upon prolonged culture in vitro. *Virology* **263**, 55–69.
- Decroly, E., Vandenbranden, M., Ruysschaert, J.M., Cogniaux, J., Jacob, G.S., Howard, S.C., Marshall, G., Kompelli, A., Basak, A., Jean, F., et al. (1994). The convertases furin and PC1 can both cleave the human immunodeficiency virus (HIV)-1 envelope glycoprotein gp160 into gp120 (HIV-1 SU) and gp41 (HIV-1 TM). *J. Biol. Chem.* **269**, 12240–12247.
- del Rio, A., Perez-Jimenez, R., Liu, R., Roca-Cusachs, P., Fernandez, J.M., and Sheetz, M.P. (2009). Stretching single talin rod molecules activates vinculin binding. *Science* **323**, 638–641.
- Dill, K.A., and Alonso, D.O.V. (1988). *Conformational Entropy and Protein Stability* (Springer Berlin Heidelberg).
- Earl, P.L., Doms, R.W., and Moss, B. (1990). Oligomeric structure of the human immunodeficiency virus type 1 envelope glycoprotein. *Proc. Natl. Acad. Sci. USA* **87**, 648–652.
- Earl, P.L., Moss, B., and Doms, R.W. (1991). Folding, interaction with GRP78-BiP, assembly, and transport of the human immunodeficiency virus type 1 envelope protein. *J. Virol.* **65**, 2047–2055.
- Eckels, E.C., Haldar, S., Tapia-Rojo, R., Rivas-Pardo, J.A., and Fernandez, J.M. (2019). The Mechanical Power of Titin Folding. *Cell Rep.* **27**, 1836–1847.e1834.
- Ellgaard, L., McCaul, N., Chatsisvili, A., and Braakman, I. (2016). Co- and Post-Translational Protein Folding in the ER. *Traffic* **17**, 615–638.
- Fenouillet, E., Papandréou, M.J., and Jones, I.M. (1997). Recombinant HIV envelope expressed in an alpha-glucosidase I-deficient CHO cell line and its parental cell line in the presence of 1-deoxynojirimycin is functional. *Virology* **231**, 89–95.
- Garces, F., Lee, J.H., de Val, N., de la Pena, A.T., Kong, L., Puchades, C., Hua, Y., Stanfield, R.L., Burton, D.R., Moore, J.P., et al. (2015). Affinity Maturation of a Potent Family of HIV Antibodies Is Primarily Focused on Accommodating or Avoiding Glycans. *Immunity* **43**, 1053–1063.
- García-Maroto, F., Castagnaro, A., Sanchez de la Hoz, P., Marañón, C., Carbonero, P., and García-Olmedo, F. (1991). Extreme variations in the ratios of non-synonymous to synonymous nucleotide substitution rates in signal peptide evolution. *FEBS Lett.* **287**, 67–70.
- Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345.
- Gogala, M., Becker, T., Beatrix, B., Armache, J.P., Barrio-Garcia, C., Berninghausen, O., and Beckmann, R. (2014). Structures of the Sec61 complex engaged in nascent peptide translocation or membrane insertion. *Nature* **506**, 107–110.
- Gordon, W.R., Zimmerman, B., He, L., Miles, L.J., Huang, J., Tianont, K., McArthur, D.G., Aster, J.C., Perrimon, N., Loparo, J.J., and Blacklow, S.C. (2015). Mechanical Allostery: Evidence for a Force Requirement in the Proteolytic Activation of Notch. *Dev. Cell* **33**, 729–736.
- Görlich, D., Hartmann, E., Prehn, S., and Rapoport, T.A. (1992a). A protein of the endoplasmic reticulum involved early in polypeptide translocation. *Nature* **357**, 47–52.
- Görlich, D., Prehn, S., Hartmann, E., Kalies, K.-U., and Rapoport, T.A. (1992b). A mammalian homolog of SEC61p and SECYp is associated with ribosomes and nascent polypeptides during translocation. *Cell* **71**, 489–503.
- Hallenberger, S., Bosch, V., Angliker, H., Shaw, E., Klenk, H.D., and Garten, W. (1992). Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. *Nature* **360**, 358–361.
- Hegde, R.S., and Bernstein, H.D. (2006). The surprising complexity of signal sequences. *Trends Biochem. Sci.* **31**, 563–571.
- Hertadi, R., Gruswitz, F., Silver, L., Koide, A., Koide, S., Arakawa, H., and Ikai, A. (2003). Unfolding mechanics of multiple OspA substructures investigated with single molecule force spectroscopy. *J. Mol. Biol.* **333**, 993–1002.
- Hoelen, H., Kleizen, B., Schmidt, A., Richardson, J., Charitou, P., Thomas, P.J., and Braakman, I. (2010). The primary folding defect and rescue of ΔF508 CFTR emerge during translation of the mutant domain. *PLoS ONE* **5**, e15458.
- Horwitz, M.S., Scharf, M.D., and Maizel, J.V., Jr. (1969). Synthesis and assembly of adenovirus 2. I. Polypeptide synthesis, assembly of capsomeres, and morphogenesis of the virion. *Virology* **39**, 682–694.

- Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802.
- Ingolia, N.T., Hussmann, J.A., and Weissman, J.S. (2019). Ribosome Profiling: Global Views of Translation. *Cold Spring Harb. Perspect. Biol.* 11, a032698.
- Jackson, R.C., and Blobel, G. (1977). Post-translational cleavage of presecretory proteins with an extract of rough microsomes from dog pancreas containing signal peptidase activity. *Proc. Natl. Acad. Sci. USA* 74, 5598–5602.
- Jansen, G., Määttä, P., Denisov, A.Y., Scarffe, L., Schade, B., Balghi, H., Dejgaard, K., Chen, L.Y., Muller, W.J., Gehring, K., and Thomas, D.Y. (2012). An interaction map of endoplasmic reticulum chaperones and foldases. *Mol. Cell. Proteomics* 11, 710–723.
- Julien, J.P., Cupo, A., Sok, D., Stanfield, R.L., Lyumkis, D., Deller, M.C., Klasse, P.J., Burton, D.R., Sanders, R.W., Moore, J.P., et al. (2013). Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science* 342, 1477–1483.
- Kanapin, A., Batalov, S., Davis, M.J., Gough, J., Grimmond, S., Kawaji, H., Magrane, M., Matsuda, H., Schönbach, C., Teasdale, R.D., and Yuan, Z. RIKEN GER Group; GSL Members (2003). Mouse proteome analysis. *Genome Res.* 13 (6B), 1335–1344.
- Knopf, P.M., and Lamfrom, H. (1965). Changes in the Ribosome Distribution during Incubation of Rabbit Reticulocytes in Vitro. *Biochim. Biophys. Acta* 95, 398–407.
- Krishna, M.M., and Englander, S.W. (2005). The N-terminal to C-terminal motif in protein folding and function. *Proc. Natl. Acad. Sci. USA* 102, 1053–1058.
- Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, J., and Hendrickson, W.A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 393, 648–659.
- Land, A., and Braakman, I. (2001). Folding of the human immunodeficiency virus type 1 envelope glycoprotein in the endoplasmic reticulum. *Biochimie* 83, 783–790.
- Land, A., Zonneveld, D., and Braakman, I. (2003). Folding of HIV-1 envelope glycoprotein involves extensive isomerization of disulfide bonds and conformation-dependent leader peptide cleavage. *FASEB J.* 17, 1058–1067.
- Lee, W.R., Syu, W.J., Du, B., Matsuda, M., Tan, S., Wolf, A., Essex, M., and Lee, T.H. (1992). Nonrandom distribution of gp120 N-linked glycosylation sites important for infectivity of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. USA* 89, 2213–2217.
- Li, Y., Luo, L., Thomas, D.Y., and Kang, C.Y. (1994). Control of expression, glycosylation, and secretion of HIV-1 gp120 by homologous and heterologous signal sequences. *Virology* 204, 266–278.
- Leonard, C.K., Spellman, M.W., Riddle, L., Harris, R.J., Thomas, J.N., and Gregory, T.J. (1990). Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J. Biol. Chem.* 265, 10373–10382.
- Li, Y., Bergeron, J.J., Luo, L., Ou, W.J., Thomas, D.Y., and Kang, C.Y. (1996). Effects of inefficient cleavage of the signal sequence of HIV-1 gp 120 on its association with calnexin, folding, and intracellular transport. *Proc. Natl. Acad. Sci. USA* 93, 9606–9611.
- Li, Y., Luo, L., Thomas, D.Y., and Kang, C.Y. (2000). The HIV-1 Env protein signal sequence retards its cleavage and down-regulates the glycoprotein folding. *Virology* 272, 417–428.
- Lingappa, V.R., Devillers-Thiery, A., and Blobel, G. (1977). Nascent pre-hormones are intermediates in the biosynthesis of authentic bovine pituitary growth hormone and prolactin. *Proc. Natl. Acad. Sci. USA* 74, 2432–2436.
- Liu, S., Cheng, W., Fowle Grider, R., Shen, G., and Li, W. (2014). Structures of an intramembrane vitamin K epoxide reductase homolog reveal control mechanisms for electron transfer. *Nat. Commun.* 5, 3110.
- Lyumkis, D., Julien, J.P., de Val, N., Cupo, A., Potter, C.S., Klasse, P.J., Burton, D.R., Sanders, R.W., Moore, J.P., Carragher, B., et al. (2013). Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science* 342, 1484–1490.
- Matczuk, A.K., Kunec, D., and Veit, M. (2013). Co-translational processing of glycoprotein 3 from equine arteritis virus: N-glycosylation adjacent to the signal peptide prevents cleavage. *J. Biol. Chem.* 288, 35396–35405.
- McCaul, N., Yeoh, H.Y., van Zadelhoff, G., Lodder, N., Kleizen, B., and Braakman, I. (2019). Analysis of Protein Folding, Transport, and Degradation in Living Cells by Radioactive Pulse Chase. *J. Vis. Exp.* (144)
- Moore, J.P., and Jarrett, R.F. (1988). Sensitive ELISA for the gp120 and gp160 surface glycoproteins of HIV-1. *AIDS Res. Hum. Retroviruses* 4, 369–379.
- Moore, P.L., Crooks, E.T., Porter, L., Zhu, P., Cayanan, C.S., Grise, H., Corcoran, P., Zwick, M.B., Franti, M., Morris, L., et al. (2006). Nature of nonfunctional envelope proteins on the surface of human immunodeficiency virus type 1. *J. Virol.* 80, 2515–2528.
- Morrison, G.M., Semple, C.A., Kilanowski, F.M., Hill, R.E., and Dorin, J.R. (2003). Signal sequence conservation and mature peptide divergence within subgroups of the murine beta-defensin gene family. *Mol. Biol. Evol.* 20, 460–470.
- Mowbray, S.L., and Koshland, D.E., Jr. (1987). Additive and independent responses in a single receptor: aspartate and maltose stimuli on the tar protein. *Cell* 50, 171–180.
- Oliver, J.D., van der Wal, F.J., Bulleid, N.J., and High, S. (1997). Interaction of the thiol-dependent reductase ERp57 with nascent glycoproteins. *Science* 275, 86–88.
- Pancera, M., Majeed, S., Ban, Y.E., Chen, L., Huang, C.C., Kong, L., Kwon, Y.D., Stuckey, J., Zhou, T., Robinson, J.E., et al. (2010). Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. *Proc. Natl. Acad. Sci. USA* 107, 1166–1171.
- Peden, K., Emerman, M., and Montagnier, L. (1991). Changes in growth properties on passage in tissue culture of viruses derived from infectious molecular clones of HIV-1LAI, HIV-1MAL, and HIV-1ELI. *Virology* 185, 661–672.
- Pfeiffer, T., Pisch, T., Devitt, G., Holtkotte, D., and Bosch, V. (2006). Effects of signal peptide exchange on HIV-1 glycoprotein expression and viral infectivity in mammalian cells. *FEBS Lett.* 580, 3775–3778.
- Piersma, D., Berns, E.M., Verhoef-Post, M., Uitterlinden, A.G., Braakman, I., Pols, H.A., and Themmen, A.P. (2006). A common polymorphism renders the luteinizing hormone receptor protein more active by improving signal peptide function and predicts adverse outcome in breast cancer patients. *J. Clin. Endocrinol. Metab.* 91, 1470–1476.
- Rehm, A., Stern, P., Ploegh, H.L., and Tortorella, D. (2001). Signal peptide cleavage of a type I membrane protein, HCMV US11, is dependent on its membrane anchor. *EMBO J.* 20, 1573–1582.
- Rognoni, L., Stigler, J., Pelz, B., Ylänne, J., and Rief, M. (2012). Dynamic force sensing of filamin revealed in single-molecule experiments. *Proc. Natl. Acad. Sci. USA* 109, 19679–19684.
- Rutkowski, D.T., Ott, C.M., Polansky, J.R., and Lingappa, V.R. (2003). Signal sequences initiate the pathway of maturation in the endoplasmic reticulum lumen. *J. Biol. Chem.* 278, 30365–30372.
- Sanders, R.W., Dankers, M.M., Busser, E., Caffrey, M., Moore, J.P., and Berkhout, B. (2004). Evolution of the HIV-1 envelope glycoproteins with a disulfide bond between gp120 and gp41. *Retrovirology* 7, 3.
- Sanders, R.W., Hsu, S.T., van Anken, E., Liscaljet, I.M., Dankers, M., Bontjer, I., Land, A., Braakman, I., Bonvin, A.M., and Berkhout, B. (2008). Evolution rescues folding of human immunodeficiency virus-1 envelope glycoprotein GP120 lacking a conserved disulfide bond. *Mol. Biol. Cell* 19, 4707–4716.
- Sauter, N.K., Hanson, J.E., Glick, G.D., Brown, J.H., Crowther, R.L., Park, S.J., Skehel, J.J., and Wiley, D.C. (1992). Binding of influenza virus hemagglutinin to analogs of its cell-surface receptor, sialic acid: analysis by proton nuclear magnetic resonance spectroscopy and X-ray crystallography. *Biochemistry* 31, 9609–9621.
- Schildknegt, D., Lodder, N., Pandey, A., Chatsisvili, A., Egmond, M., Pena, F., Braakman, I., and van der Sluijs, P. (2019). Characterization of CNPY5 and its family members. *Protein Sci.* 28, 1276–1289.



- Schulman, S., Wang, B., Li, W., and Rapoport, T.A. (2010). Vitamin K epoxide reductase prefers ER membrane-anchored thioredoxin-like redox partners. *Proc. Natl. Acad. Sci. USA* *107*, 15027–15032.
- Seidel, E., Dassa, L., Kahlon, S., Tirosh, B., Helenius, A., Seidel Malkinson, T., and Mandelboim, O. (2021). A slowly cleaved viral signal peptide acts as a protein-integral immune evasion domain. *Nat. Commun.* *12*, 2061.
- Snapp, E.L., McCaul, N., Quandt, M., Cabartova, Z., Bontjer, I., Källgren, C., Nilsson, I., Land, A., von Heijne, G., Sanders, R.W., and Braakman, I. (2017). Structure and topology around the cleavage site regulate post-translational cleavage of the HIV-1 gp160 signal peptide. *eLife* *6*, e26067.
- Soler, M.A., and Faisca, P.F. (2012). How difficult is it to fold a knotted protein? In silico insights from surface-tethered folding experiments. *PLoS ONE* *7*, e52343.
- Srinivasan, N., Sowdhamini, R., Ramakrishnan, C., and Balaram, P. (1990). Conformations of disulfide bridges in proteins. *Int. J. Pept. Protein Res.* *36*, 147–155.
- Sun, X., Li, Q., Wu, Y., Wang, M., Liu, Y., Qi, J., Vavricka, C.J., and Gao, G.F. (2014). Structure of influenza virus N7: the last piece of the neuraminidase “jigsaw” puzzle. *J. Virol.* *88*, 9197–9207.
- Tamura, T., Cormier, J.H., and Hebert, D.N. (2011). Characterization of early EDEM1 protein maturation events and their functional implications. *J. Biol. Chem.* *286*, 24906–24915.
- Tatu, U., Braakman, I., and Helenius, A. (1993). Membrane glycoprotein folding, oligomerization and intracellular transport: effects of dithiothreitol in living cells. *EMBO J.* *12*, 2151–2157.
- Tatu, U., Hammond, C., and Helenius, A. (1995). Folding and oligomerization of influenza hemagglutinin in the ER and the intermediate compartment. *EMBO J.* *14*, 1340–1348.
- van Anken, E., Sanders, R.W., Liscaljet, I.M., Land, A., Bontjer, I., Tillemans, S., Nabatov, A.A., Paxton, W.A., Berkhout, B., and Braakman, I. (2008). Only five of 10 strictly conserved disulfide bonds are essential for folding and eight for function of the HIV-1 envelope glycoprotein. *Mol. Biol. Cell* *19*, 4298–4309.
- Van Damme, N., Goff, D., Katsura, C., Jorgenson, R.L., Mitchell, R., Johnson, M.C., Stephens, E.B., and Guatelli, J. (2008). The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein. *Cell Host Microbe* *3*, 245–252.
- van Gils, M.J., van den Kerkhof, T.L., Ozorowski, G., Cottrell, C.A., Sok, D., Pauthner, M., Pallesen, J., de Val, N., Yasmeen, A., de Taeye, S.W., et al. (2016). An HIV-1 antibody from an elite neutralizer implicates the fusion peptide as a site of vulnerability. *Nat. Microbiol.* *2*, 16199.
- Veitia, R.A., and Caburet, S. (2009). Extensive sequence turnover of the signal peptides of members of the GDF/BMP family: exploring their evolutionary landscape. *Biol. Direct* *4*, 22.
- von Heijne, G. (1983). Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem.* *133*, 17–21.
- von Heijne, G. (1984). Analysis of the distribution of charged residues in the N-terminal region of signal sequences: implications for protein export in prokaryotic and eukaryotic cells. *EMBO J.* *3*, 2315–2318.
- von Heijne, G. (1985). Signal sequences. The limits of variation. *J. Mol. Biol.* *184*, 99–105.
- Walter, P., and Blobel, G. (1981). Translocation of proteins across the endoplasmic reticulum III. Signal recognition protein (SRP) causes signal sequence-dependent and site-specific arrest of chain elongation that is released by microsomal membranes. *J. Cell Biol.* *91*, 557–561.
- Wei, X., Decker, J.M., Liu, H., Zhang, Z., Arani, R.B., Kilby, J.M., Saag, M.S., Wu, X., Shaw, G.M., and Kappes, J.C. (2002). Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. *Antimicrob. Agents Chemother.* *46*, 1896–1905.
- Weissman, J.S., and Kim, P.S. (1992). The pro region of BPTI facilitates folding. *Cell* *71*, 841–851.
- Weissman, J.S., and Kim, P.S. (1993). Efficient catalysis of disulphide bond rearrangements by protein disulphide isomerase. *Nature* *365*, 185–188.
- Weissman, J.S., and Kim, P.S. (1995). A kinetic explanation for the rearrangement pathway of BPTI folding. *Nat. Struct. Biol.* *2*, 1123–1130.
- Williams, E.J., Pal, C., and Hurst, L.D. (2000). The molecular evolution of signal peptides. *Gene* *253*, 313–322.
- Wyatt, R., and Sodroski, J. (1998). The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science* *280*, 1884–1888.
- Yang, X., Mahony, E., Holm, G.H., Kassa, A., and Sodroski, J. (2003). Role of the gp120 inner domain  $\beta$ -sandwich in the interaction between the human immunodeficiency virus envelope glycoprotein subunits. *Virology* *313*, 117–125.
- Yao, M., Goult, B.T., Chen, H., Cong, P., Sheetz, M.P., and Yan, J. (2014). Mechanical activation of vinculin binding to talin locks talin in an unfolded conformation. *Sci. Rep.* *4*, 4610.
- Zhang, X., Halvorsen, K., Zhang, C.Z., Wong, W.P., and Springer, T.A. (2009). Mechanoenzymatic cleavage of the ultralarge vascular protein von Willebrand factor. *Science* *324*, 1330–1334.
- Zhou, H.X. (2008). Protein folding in confined and crowded environments. *Arch. Biochem. Biophys.* *469*, 76–82.
- Zhou, H.X., and Dill, K.A. (2001). Stabilization of proteins in confined spaces. *Biochemistry* *40*, 11289–11293.
- Zschenker, O., Jung, N., Rethmeier, J., Trautwein, S., Hertel, S., Zeigler, M., and Ameis, D. (2001). Characterization of lysosomal acid lipase mutations in the signal peptide and mature polypeptide region causing Wolman disease. *J. Lipid Res.* *42*, 1033–1040.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Polyclonal Rabbit anti-gp160	Braakman lab ( <a href="#">Land and Braakman, 2001</a> )	40336
Polyclonal goat anti-gp120	NIH HIV Reagent program	HT3, NIH531
Polyclonal Rabbit anti-HA tag	Braakman lab ( <a href="#">Schildknegt et al., 2019</a> )	MrBrown
<b>Chemicals, peptides, and recombinant proteins</b>		
Thrombin	Sigma Aldrich	T-6634
<b>Experimental models: Cell lines</b>		
TZM-bl	NIH HIV Reagent program	Cat#8129–442, RRID:CVCL_B478
HEK293T	ATCC	CRL-11268
HeLa	ATCC	CCL-2
C33A	ATCC	HTB-31
<b>Recombinant DNA</b>		
pMQ	Braakman lab ( <a href="#">Snapp et al., 2017</a> )	N/A
pRS1	Sanders lab ( <a href="#">Sanders et al., 2004</a> )	N/A
pLAI	NIH HIV Reagent Program	ARP-2532
pSV7d	Quinnan lab ( <a href="#">Cham et al., 2006</a> )	N/A
pSG3 <sup>ΔEnv</sup>	NIH HIV Reagent Program	ARP-11051
<b>Software and algorithms</b>		
Prism v9	Graphpad	RRID:SCR_002798
ImageQuant-TL	Cytiva	RRID:SCR_018374

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Ineke Braakman ([i.braakman@uu.nl](mailto:i.braakman@uu.nl)).

### Materials availability

All requests for resources and reagents should be directed to the Lead Contact author. This includes antibodies, plasmids and proteins. All reagents are available on request after completion of a Material Transfer Agreement, but we may require a payment if there is potential for commercial application.

### Data and code availability

This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request. All data reported in this paper will be shared by the lead contact upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Cell lines

The TZM-bl reporter cell line (RRID: CVCL\_B478), obtained from NIH HIV Reagent Program, Division of AIDS, NIAID, NIH (Dr. John C. Kappes and Dr. Xiaoyun Wu), the HEK293T cell line (ATCC, CRL-11268, RRID:CVCL\_1926) were and the C33A (ATCC, HTB-31, RRID: CVCL\_1094) cell line were cultured in Dulbecco's modified Eagle medium (GIBCO) containing 10% FCS, 100 units/mL penicillin and 100 μg/mL streptomycin.

HeLa cells (ATCC, CCL-2, RRID: CVCL\_0030) were maintained in MEM containing 10% FCS, nonessential amino acids, glutamax and penicillin/streptomycin (100 U/mL). All cell lines were cultured at 37°C with 5% CO<sub>2</sub> and routinely tested negative for mycoplasma contamination. All of the cell lines used in this study are female and derived from humans. All cell lines were assumedly authenticated by their representative sources and were not further authenticated for this study.

## Viruses

HIV-1<sub>LAI</sub> was used in this study, propagated from the full-length molecular clone pLAI (NIH ARP-2532). Pseudoviruses using the HIV<sub>SG3</sub> backbone were generated using the plasmid pSG3<sup>ΔEnv</sup> (NIH HIV Reagent Program, Division of AIDS, NIAID, NIH: ARP-11051, contributed by Drs. John C. Jappes and Xiaoyun Wu).

Although we studied gp160 of the LAI isolate, we followed the canonical HXB2 residue numbering (GenBank: K03455.1), which relates to the LAI numbering as follows: because of an insertion of five residues in the V1 loop of LAI gp160, all cysteine residues beyond this loop have a number that is 5 residues lower in HXB2 than in LAI: until Cys131, numbering is identical, but Cys162 in LAI becomes 157 in HXB2, etc (Figure S7).

## METHOD DETAILS

### Plasmids, antibodies and reagents

The full-length molecular clone of HIV-1<sub>LAI</sub> (pLAI) was the source of wild-type and mutant viruses (Peden et al., 1991). The QuikChange Site-Directed Mutagenesis kit (Stratagene) was used to introduce mutations into *Env* in plasmid pRS1 as described before; the entire *Env* gene was verified by DNA sequencing (Sanders et al., 2004). Mutant *Env* genes from pRS1 were cloned back into pLAI as Sall-BamHI fragments. The plasmids pcDNA3 LAI gp160 and pSV7d JR-FL gp160 were used for making wild-type and mutant pseudo-viruses. All point mutations were introduced using QuikChange Site-Directed mutagenesis as above.

For transient transfection of gp120/160 we used the previously described pMQ plasmid (Snapp et al., 2017). C-terminal truncations were generated by PCR of wt gp120 and Gibson assembled back into XbaI/XhoI digested pMQ. The thrombin-cleavable construct was designed based on stable V1V2 loop deletion number 2 (Bontjer et al., 2009) and generated from gp120 C119-205A using Gibson assembly (Gibson et al., 2009). All point mutations were introduced using QuikChange Site-Directed mutagenesis as above.

For immunoprecipitation: we used the previously described polyclonal rabbit anti-gp160 antibody 40336 which recognizes all forms of gp120 (Land et al., 2003), polyclonal antibody HT3 (NIH531) which was obtained from the NIH AIDS reagent program and  $\alpha$ -HA tag antibody “MrBrown” produced by us (Schildknegt et al., 2019).

Thrombin was purchased as a lyophilized powder from Sigma Aldrich (T-6634) and stored in thrombin-storage buffer [50 mM Sodium Citrate pH 6.5, 200 mM NaCl, 0.1% BSA (w/v), 50% glycerol (w/v)].

### Transient transfection

Twenty-four hours before pulse labeling, HeLa cells were transfected with pMQ gp120/gp160 or HA constructs using polyethylenimine (Polysciences) as described before (Hoelen et al., 2010).

### Virus production

HIV-1 LAI virus stocks were produced by transfecting HEK293T cells with wild-type or mutant pLAI constructs using the Lipofectamine 2000 Transfection Reagent (Invitrogen) per the manufacturer's protocol. HIV-1 LAI pseudovirus stocks were produced by co-transfecting C33A cells with wild-type or mutant pcDNA3 LAI gp160 constructs in a 2:1 ratio with the pSG3<sup>ΔEnv</sup> (Env-deleted) backbone construct using the calcium-phosphate precipitation as described previously (Das et al., 1999). HIV-1 JR-FL pseudovirus stocks were produced by co-transfecting HEK293T cells with wild-type or mutant pSV7d JR-FL gp160 constructs in a 2:1 ratio with pSG3<sup>ΔEnv</sup> (see above) using the Lipofectamine 2000 method as above. The virus-containing culture supernatants were harvested 2 days post-transfection, stored at  $-80^{\circ}\text{C}$ , and the (pseudo-) virus concentrations were quantitated by CA-p24 ELISA as described previously (Moore and Jarrett, 1988). These values were used to normalize the amount of virus used in subsequent infection experiments.

Virus-like particles (VLPs) were prepared as previously reported (Crooks et al., 2011). In brief, HEK293Ts were co-transfected with Env-expressing plasmid and the Env-deficient HIV-1 genomic backbone plasmid pNL-LucR-E<sup>-</sup>. All Env mutants were made in the pCAGGS vector in JR-FL SOS E168K gp160. VLPs were purified as previously described (Moore et al., 2006).

### Infectivity assay

The TZM-bl reporter cell line stably expresses high levels of CD4 and HIV-1 coreceptors CCR5 and CXCR4 and contains the luciferase and  $\beta$ -galactosidase genes under the control of the HIV-1 long-terminal-repeat (LTR) promoter (Wei et al., 2002). Single-cycle infectivity assays were performed as described before (Bontjer et al., 2009, 2010; Snapp et al., 2017). In short, one day prior to infection,  $17 \times 10^3$  TZM-bl cells per well were plated on a 96-well plate in DMEM containing 10% FCS, 100 units/mL penicillin and 100  $\mu\text{g}/\text{mL}$  streptomycin and incubated at  $37^{\circ}\text{C}$  with 5%  $\text{CO}_2$ . A fixed amount of HIV-1 LAI virus (500 pg of CA-p24) or a fixed amount of HIV-1 JR-FL or HIV-1 LAI pseudovirus (1,000 pg of CA-p24) was added to the cells (70%–80% confluency) in the presence of 400 nM saquinavir (Roche) to block secondary rounds of infection and 40  $\mu\text{g}/\text{mL}$  DEAE in a total volume of 200  $\mu\text{L}$ . Two days post-infection, medium was removed, cells were washed with phosphate-buffered saline (50 mM sodium phosphate buffer, pH 7.0, 150 mM NaCl) and lysed in Reporter Lysis buffer (Promega). Luciferase activity was measured using a Luciferase Assay kit (Promega) and a Glomax luminometer (Turner BioSystems) per manufacturer's instructions. Uninfected cells were used to correct

for background luciferase activity. All infections were performed in quadruplicate. Titrations were performed with each virus batch at two different amounts of virus (Figure S6).

### Folding assay

HeLa cells transfected with wild-type or mutant gp160/gp120 constructs were subjected to pulse-chase analysis as described before (Land et al., 2003; McCaul et al., 2019; Snapp et al., 2017). In short, cells were starved for cysteine and methionine for 15–30 min and pulse labeled for 5 min with 55  $\mu\text{Ci}$ /35-mm dish of Easytag express  $^{35}\text{S}$  protein labeling mix (Perkin Elmer). Where indicated (+DTT), cells were incubated with 5 mM DTT for 5 min before and during the pulse. The pulse was stopped, and chase started by the first of 2 washes with chase medium containing an excess of unlabeled cysteine and methionine. At the end of each chase, medium was collected, and cells were cooled on ice and further disulfide bond formation and isomerization was blocked with 20 mM iodoacetamide. Cells were lysed and detergent lysates and medium samples were subjected to overnight immunoprecipitation at 4°C with polyclonal antibody 40336 against gp160.

### Deglycosylation, SDS-PAGE, and autoradiography

Where appropriate, to identify gp160 folding intermediates, glycans were removed from lysate-derived gp120 or gp160 with Endoglycosidase H (Roche) treatment of the immunoprecipitates as described before (Land et al., 2003). This decreases electrophoretic mobility and thereby increases mobility differences between folding intermediates and native gp120 or gp160.

Samples were subjected to non-reducing and reducing (25 mM DTT) SDS-PAGE. Gels were dried and exposed to super-resolution phosphor screens (FujiFilm) or Kodak Biomax MR films (Carestream). Phosphor screens were scanned with a Typhoon FLA-7000 scanner (GE Healthcare Life Sciences).

### mPEG treatment

HEK293T cells transfected with wild-type or mutant 111X were subjected to radioactive labeling as described above. At the end of the labeling, cells were transferred to ice and incubated in Dulbecco's PBS without  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  containing 20 mM *N*-ethyl maleimide (NEM) and 5 mM EDTA. Cells then were subjected to a modified "double-alkylation variant" mPEG treatment as described by Appenzeller-Herzog and Ellgaard (Appenzeller-Herzog and Ellgaard, 2008). In short, cells were homogenized by passage through a 25-G needle and proteins denatured with 2% SDS for 1 h @ 95°C. Samples then were alkylated again with 20 mM NEM before immunoprecipitation with anti-HA tag antibody MrBrown for 2 hours at 4°C. After immunoprecipitation, samples were denatured and reduced with 25 mM tris(2-carboxylethyl)phosphine (TCEP) followed by incubation with 15 mM mPEG-maleimide 5000 for 1 h at room temperature. Samples were immunoprecipitated again via the HA-tag and analyzed by 4%–15% non-reducing gradient SDS-PAGE (BioRad) and processed as before.

### Thrombin cleavage

After HeLa cells transfected with various thrombin-cleavable constructs were pulse-labeled as described above, detergent lysates were immunoprecipitated with antibody HT3 for 1 h at 4°C with rotation. Immunoprecipitates were washed and resuspended in 15  $\mu\text{l}$  thrombin cleavage buffer (20 mM Tris-HCl, pH 8.4, 150 mM NaCl, 2.5 mM  $\text{CaCl}_2$ ) + 0.2% SDS and denatured for 5 minutes at 95°C. SDS was quenched by addition of 10  $\mu\text{l}$  cleavage buffer + 2% Tx100. Thrombin (0.75 U) in 5  $\mu\text{l}$  cleavage buffer then was added to samples and incubated for exactly 15 minutes. For mock-digested samples, an equivalent volume of thrombin storage buffer was added instead. Digestion was stopped by the addition of hot (95°C) 5X sample buffer and immediately placing in a 95°C heat block for 5 minutes. Samples then were subjected to non-reducing or reducing (25 mM DTT) 15%–20% discontinuous-gradient SDS-PAGE and processed as before.

### Native PAGE and western blotting

Blue native PAGE (BN-PAGE) was performed as described previously (Crooks et al., 2011). Briefly, VLPs were washed and gently solubilized in 0.12% Triton X-100 in 1 mM EDTA-1.5 M aminocaproic acid with a protease inhibitor cocktail containing 4-(2-aminoethyl)benzenesulfonyl fluoride, E-64, bestatin, leupeptin, aprotinin, and sodium EDTA (P-2714; Sigma). For cell fraction, cells were homogenized in PBS + protease inhibitors, followed by addition of 1% Triton X-100. Lysate was then purified over Gallanthus Nivalis lectin agarose. An equal volume of 2x sample buffer (100 mM morpholinepropanesulfonic acid (MOPS), 100 mM Tris-HCl, pH 7.7, 40% glycerol, and 0.1% Coomassie blue) was added. Samples were loaded onto a 4%–12% Bis-Tris NuPAGE gel (ThermoFisher) and separated at 4°C for 3h at 100 V. Ferritin (Amersham) was used as a size standard. The gel was then transferred to PVDF, destained and blocked (4% fat-free milk in PBS), and probed with an anti-gp120 (MAbs 2G12, b12, E51, and 39F at 1  $\mu\text{g}/\text{ml}$  each). Blots were then probed by anti-human Fc alkaline phosphatase conjugate (Jackson) and developed using SigmaFast BCIP/NBT substrate (Sigma).

### QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification of phosphor screens from pulse chase experiments were performed with ImageQuantTL software (GE Healthcare Life Sciences). Statistics for each experiment were calculated using Prism 9 (Graphpad). For experiments in Figures 2D–2F differences

were assessed using a one-way ANOVA with follow-up testing to analyze differences between specific pairs with p values corrected for multiple comparisons. For experiments in [Figures 3F](#) and [4D](#) differences were assessed using paired t tests between wild-type and mutants. For experiments in [Figures 6A–6D](#) differences were assessed using unpaired t tests. A complete list of all pairs examined, statistical methods and resulting p values can be found in Table S1. N value for pulse chase experiments ([Figures 2D](#), [2E](#), and [4D](#)) indicates number of replicate experiments performed with fresh transfectants. For virus experiments ([Figures 2F](#) and [6](#)) represent number of individual samples analyzed in each assay.