UCSF
UC San Francisco Previously Published Works

Title
Highly Mutable Linker Regions Regulate HIV-1 Rev Function and Stability

Permalink
https://escholarship.org/uc/item/6fb6x7vs

Journal
Scientific Reports, 9(1)

ISSN
2045-2322

Authors
Jayaraman, Bhargavi
Fernandes, Jason D
Yang, Shumin
et al.

Publication Date
2019

DOI
10.1038/s41598-019-41582-7

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

SCIENTIFIC REPORTS

OPEN

Highly Mutable Linker Regions Regulate HIV-1 Rev Function and Stability

Received: 22 October 2018
Accepted: 5 March 2019
Published online: 26 March 2019

Bhargavi Jayaraman[1], Jason D. Fernandes[1,3], Shumin Yang[1,2], Cynthia Smith[1] & Alan D. Frankel[1]

HIV-1 Rev is an essential viral regulatory protein that facilitates the nuclear export of intron-containing viral mRNAs. It is organized into structured, functionally well-characterized motifs joined by less understood linker regions. Our recent competitive deep mutational scanning study confirmed many known constraints in Rev's established motifs, but also identified positions of mutational plasticity, most notably in surrounding linker regions. Here, we probe the mutational limits of these linkers by testing the activities of multiple truncation and mass substitution mutations. We find that these regions possess previously unknown structural, functional or regulatory roles, not apparent from systematic point mutational approaches. Specifically, the N- and C-termini of Rev contribute to protein stability; mutations in a turn that connects the two main helices of Rev have different effects in different contexts; and a linker region which connects the second helix of Rev to its nuclear export sequence has structural requirements for function. Thus, Rev function extends beyond its characterized motifs, and is tuned by determinants within seemingly plastic portions of its sequence. Additionally, Rev's ability to tolerate many of these massive truncations and substitutions illustrates the overall mutational and functional robustness inherent in this viral protein.

Proteins balance optimal functionality with mutational tolerance in order to adapt to changes in selection pressures[1]. Proteins with high tolerance for mutation are considered genetically "robust" or "plastic", while proteins with a low mutability are considered genetically "fragile" or "brittle"[2]. The existence of a protein sequence as fragile or robust is a result of evolutionary pressures acting on the protein, such as mutation rate and biological function. Fragility and robustness are generally expected to couple with protein structure and disorder: disordered regions are expected to be robust while structured regions are comparatively brittle[3,4].

Viral proteins maintain their function amidst many forces such as high viral polymerase error rates, immune pressures, and even competing selective pressures from overlapping reading frames and RNA structures. HIV-1 Rev is one such essential viral protein which facilitates the nuclear export of intron-containing viral RNAs that encode essential viral structural and enzymatic proteins and provide full length genomes for encapsidation[5–7]. Rev experiences a high mutation rate in HIV (~10^{-3} mutations per base per cell[8]) and is also overlapped with two other essential viral proteins, Tat and Env (Fig. 1A). Overlap between viral genes is common[9], but most HIV-1 genes have at least one region with no overlap in which they can encode critical functional domains in an unconstrained way (as we recently observed with HIV-1 *tat*[10]). In contrast, *rev* shares coding information with other viral genes throughout its length (the only HIV-1 gene encoded in such a manner) meaning that its evolution is coupled to these overlapping genes.

Because of the overlap, the true functional importance of each Rev residue is difficult to assess from traditional sequence conservation analyses and instead requires careful mutational dissection in non-overlapped contexts. To address this, we recently performed competitive deep mutational scanning (CDMS) in non-overlapped viral replication assays[10]. These experiments allowed us to examine for the first time, the amino acid preferences/fitness in Rev at each position, when unconstrained by overlapping genes.

[1]Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, CA, 94158, USA. [2]School of Medicine, Tsinghua University, Beijing, 100084, China. [3]Present address: UCSC Genomics Institute/Howard Hughes Medical Institute, University of Santa Cruz, Santa Cruz, CA, 95060, USA. Bhargavi Jayaraman and Jason D. Fernandes contributed equally. Correspondence and requests for materials should be addressed to A.D.F. (email: frankel@cgl.ucsf.edu)

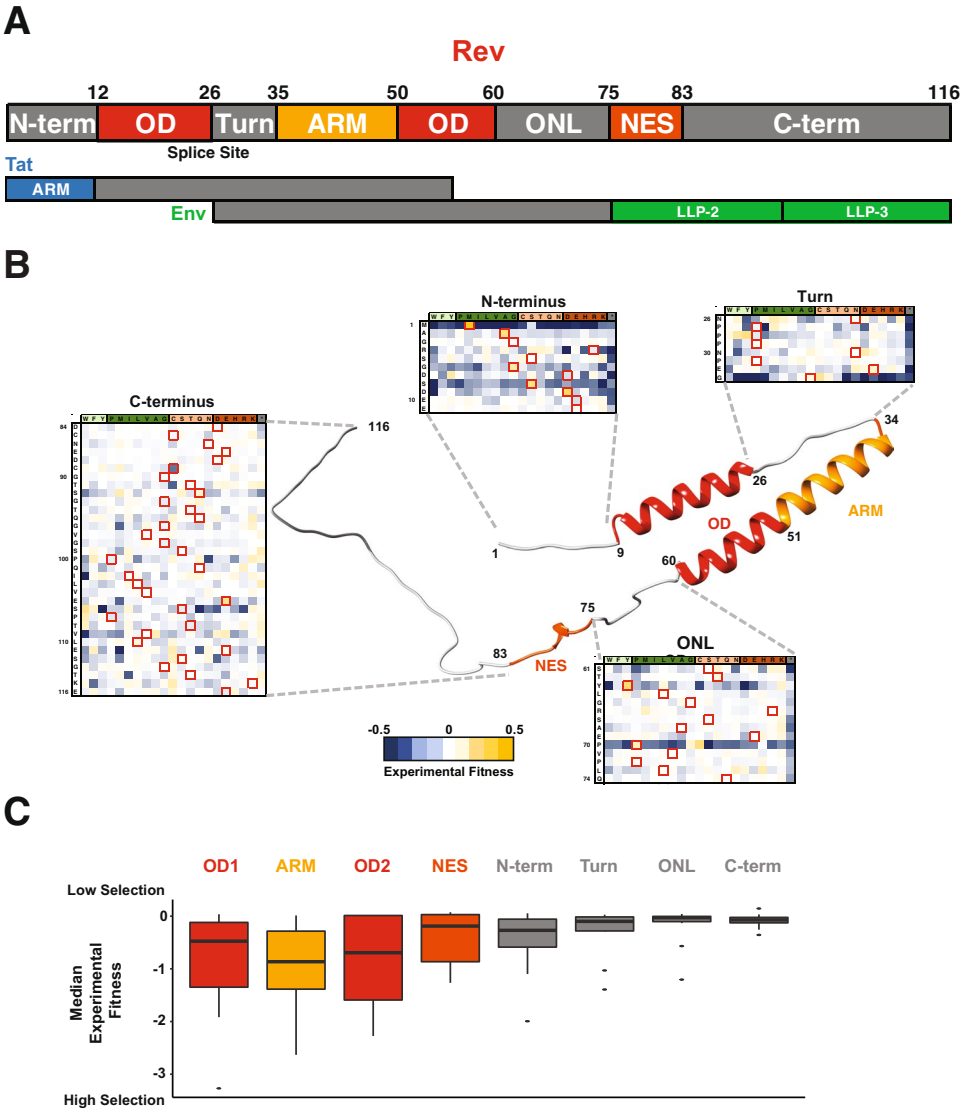

Figure 1. Structure and Organization of HIV-1 Rev. (**A**) Domain organization of Rev protein (NL4-3/HXB2 numbering) with interaction surfaces in shades of red (OD: Oligomerization domain; ARM: Arginine-Rich Motif; NES: Nuclear Export Sequence) and linker regions in grey (N-term: N-terminus, Turn, ONL: OD-NES Linker). In the viral genome, the Rev coding sequence consists of two exons (Fig. 1A; "splice site" denotes exon-exon junction as it would occur in the mRNA) both of which are contained entirely within (alternative reading frame) exons of two other HIV genes, Tat and Env. Putative structured domains of Tat (ARM: Arginine Rich Motif) and Env (LLP: Lentivirus lytic peptide) are shown along with their overlap with Rev domains. (**B**) Structural model of Rev built using Rev crystal structures (residues 8–62 from PDBID: 3LPH; residues 76–83 from PDBID: 3NBZ; all other regions were built in PyMOL for visualization purposes). Also shown is the experimental fitness of every residue in the Rev linker regions (from[10]) (**C**). Distribution of median experimental fitness values of each residue for different regions of Rev classified by domain organization (data from[10]). Specifically, there are 116 total data points each representing a single residue of Rev, each binned into the appropriate domain; the representative value for each residue is the median fitness of all 21 side chains (and the stop codon) at that position).

Results from these experiments showed that, in general, the known structured/functional regions of Rev experience selective pressure for specific side chains while the linker regions between/flanking these structured domains appear to be relatively "plastic", freely mutating between many different side chains with a variety of chemical properties (Fig. 1B,C). However, the true extent of this plasticity is unclear, as the approach only examined single point mutations within a reference viral strain background.

In this study, we expand our analyses of the point mutational datasets to evaluate the limits of the structural and mutational plasticity of Rev's structured domains and unstructured linker regions by designing deletion and mass substitution mutations and testing them for function in reporter assays and viral replication assays. Our results indicate that most of the linker regions possess previously unappreciated structural, functional or regulatory roles, not apparent from systematic point mutational approaches, while the structured regions tolerate

more variation than expected. The N- and C-termini control Rev stability; mutations in a turn that connects the two main helices of Rev have different effects in nuclear export assays and viral replication assays, suggesting an alternate functional role besides nuclear export for this region; and a linker region connecting the second helix of Rev to its nuclear export sequence becomes structured at some point in Rev's export function. Overall, our study illustrates for the first time how these linker regions are closely involved in Rev function either by direct functional roles or by more subtle regulatory roles. In sum, these experiments make evident Rev's extraordinary genetic robustness, which is achieved via the use of malleable macromolecular interaction surfaces spaced by even more malleable linker regions that can more finely tune those interactions.

Results

Competitive deep mutational scanning (CDMS) data recapitulates Rev domain structure. Rev plays an essential role in the viral life cycle, coordinating the nuclear export of unspliced or partially spliced mRNAs for translation of key viral proteins or encapsidation of full-length viral genomes. Rev recognizes a ~350-nt structured RNA within the intron known as the Rev Response Element (RRE) using an arginine-rich motif (ARM) for binding, and forms an oligomeric assembly on the RRE using two hydrophobic oligomerization motifs or domains (ODs)[11,12]. This Rev-RRE complex then recruits the host nuclear export factor Crm1, using a nuclear export sequence (NES) in Rev, together with RanGTP to transport these RNAs to the cytoplasm and thereby circumvent splicing[5,6].

The three main functional regions of Rev (ARM, OD, and NES) are flanked by linker regions[13–17] (Fig. 1A,B). The structure of the OD is formed by the interaction of two OD helices, where the second helix extends beyond the 15 residues of the helical ARM[15]. The NES forms its structure upon binding to Crm1[18]. The regions that flank these functional domains include: a short N-terminal tail (N-term), a proline-rich turn between the first and second helices (turn), a linker between the second helix and the NES (OD-NES Linker: ONL) and a long, disordered C-terminal tail (C-term)[6,19] (Fig. 1B).

We utilized the fitness values for every possible Rev allele derived from a competitive deep mutational scanning (CDMS) dataset, where Rev was engineered into non-overlapped viruses and all substitutions at individual positions were competitively replicated, to deduce functional signatures of amino acid selection[10]. These data are consistent with our current biochemical and structural understanding of Rev, as we observe that the structured regions display low mutational tolerance (low median fitness in these regions indicates that only a few side chains are fit at each position) whereas the linker regions display high mutational tolerance (near-neutral median fitness indicates that almost all side chains are fit) (Fig. 1B,C). The mutational plasticity of the linker regions is also apparent in the heat maps of all individual point mutations, further corroborating the underlying structural and functional boundaries (Fig. 1B).

The Rev OD is highly plastic at the sequence and structural levels. We first examined the amino acid requirements of the OD in more detail. Rev function requires binding of a homo-oligomer of 6–10 subunits to the RRE[11,12,19–22]. The OD is comprised of two helices which contain hydrophobic residues necessary both for oligomerization and for interacting with host proteins, such as hnRNPs and RNA helicases[23–26]. The OD forms three hydrophobic surfaces with some residues shared between the surfaces: 1) a monomer stabilization interface in which side chains from each helix of the OD (residues 19, 22, 52 and 59) stabilize the monomer structure of Rev, 2) a dimer interface which supports formation of a Rev dimer upon nucleation of Rev assembly on the RRE (residues 18, 22, 55 and 59 and residue 21 in one dimer configuration), and 3) a higher-order interface opposite the dimer interface which allows the formation of the larger Rev oligomer (residues 12, 16 and 60)[13,14,27,28] (Fig. 2A).

One very provocative feature of the Rev OD is that variation in the packing of its hydrophobic residues leads to multiple configurations of Rev dimers and oligomers[15,29,30]. This structural plasticity underlies the ability of Rev to assemble into an oligomeric complex on the RRE and raises the question of how constrained these side chains must be. CDMS data[10] show that interaction residues in each interface display a general preference for hydrophobicity and size with occasional preference for aliphatic side chains over aromatic ones (Fig. 2B). For instance, in the dimer interface, residues 22, 55, and 59 prefer large aliphatic side chains while aromatic substitutions are equally as fit as aliphatic ones at residue 19. Similar patterns are seen in the higher-order oligomerization interface with residues 12 and 60 accepting any large hydrophobic, residues 16, 55 and 60 having distinct preferences for aliphatic residues. Interestingly, residue 23, although not part of currently known oligomeric interface configurations, strongly prefers F or Y (Fig. 2B). In summary, this ability of the OD to make functionally neutral mutations between chemically similar side chains (e.g. I, L & V and in some cases W, F, and Y) at such functional critical positions increases the mutational robustness and prevents the "brittleness" that essential interfaces in other proteins can possess.

The Rev NES and ARM balance fragility and plasticity. The Rev ARM serves as both the RRE-binding domain as well as a nuclear localization signal (NLS) through interactions with Importin-β[31]. It can recognize diverse RNA sites depending on the RNA structure and the orientation/placement of ARM residues[15,20,32]. Additionally, residues in the ARM also enhance the monomer core structure through hydrogen bonds[13]. Given these numerous functional constraints, most residues of the ARM, including the non-basic residues Q36 and W45, show strong selection in the CDMS study (Fig. 3A). Another residue, Q51, implicated in stabilizing an RNA-bound Rev dimer through hydrogen bonds[15] also shows strong selection, reaffirming its importance in Rev-RRE assembly.

The Rev NES recruits Crm1 to the Rev-RRE RNP, and recent EM studies show that a Crm1 dimer is associated with the RNP[33]. Structural studies of a Rev NES peptide bound to Crm1 have shown that Rev binds Crm1 in a non-canonical manner (Fig. 3B), using prolines to properly space hydrophobic residues to contact Crm1.

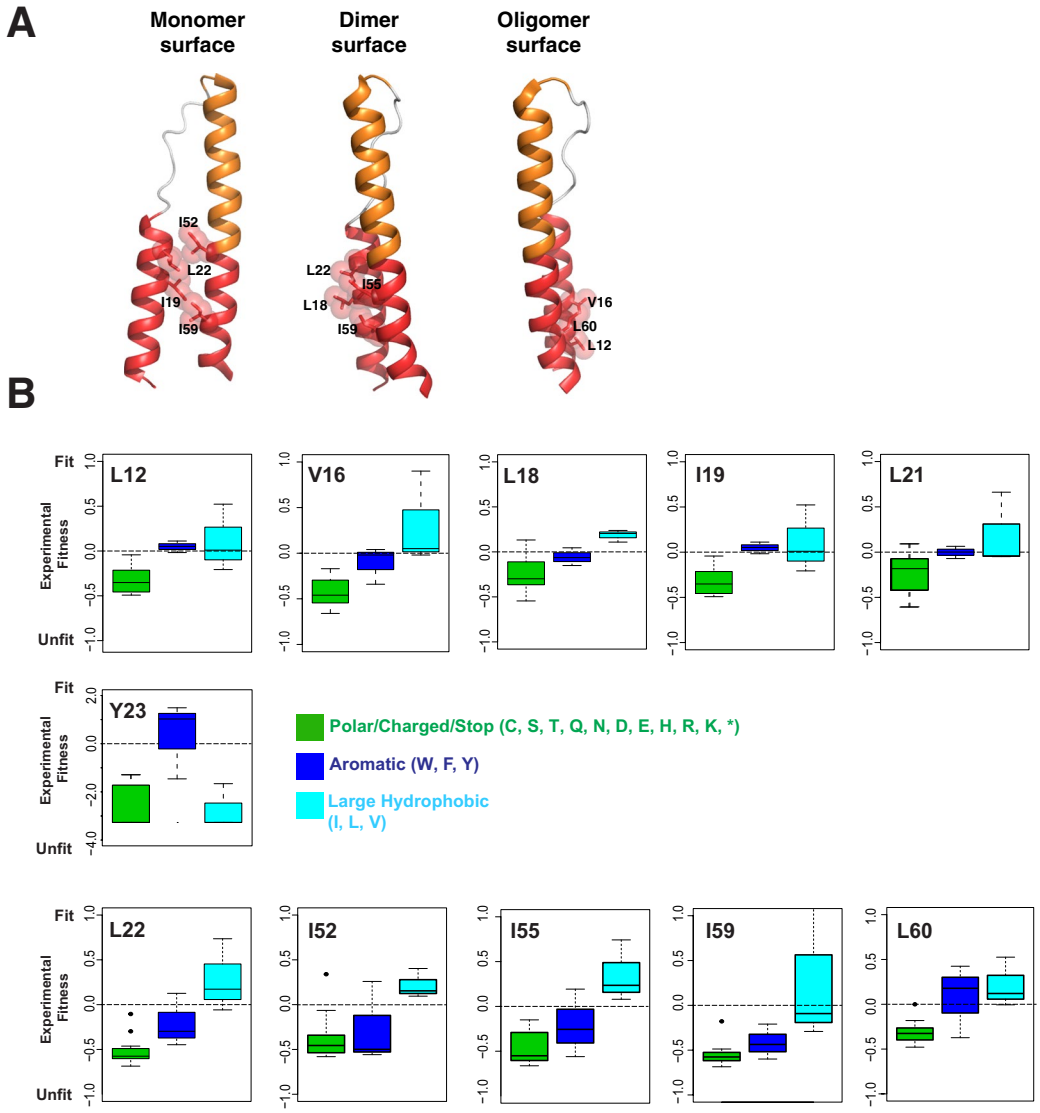

Figure 2. Selection profile of hydrophobic residues for the Rev Oligomerization Domain (OD) in the CDMS dataset[10]. (**A**) Residues from the Rev OD involved in monomer, dimer or oligomer interface packing. Structures from PDBID 3LPH and 2 × 7 L were used to generate the images. (**B**) Amino acid preferences for important Rev OD contact residues using CDMS fitness values[10]. All residues display selection against polar/charged (C,S,T,Q,N,D, E, H, R, K) residues (and stop codons) (green) while some prefer large, hydrophobic side chains (teal) over aromatics (blue). Note change of scaling for Y23.

Curiously, CDMS data shows that P76, which contacts the Crm1 binding groove, is as fit as a hydrophobic residue (L, I), whereas L75 prefers to be a proline (Fig. 3C). Furthermore, like the OD, the hydrophobic residues in the NES that contact Crm1 are tolerant of most large aliphatic side chains (Fig. 3C,D).

In order to further evaluate the effect of Rev mutations within both the structured domains and linkers, we conducted a series of RNA export reporter assays together with replication assays of non-overlapped mutant viruses to avoid the confounding effects of overlapping reading frames (Fig. S1)[10]. The reporter assay monitors Rev-dependent p24 capsid expression from a CMV promoter-driven truncated viral genome in 293 T cells[34,35] and the replication assay utilizes *rev-in-nef* viruses as previously described[10]. The combination of assays also allows us to potentially uncover additional functions of Rev in the replication context.

We tested both Rev L75P and P76L mutants for function in our reporter assay and observed only 60% activity (Fig. 3E), while the corresponding viral spreads were comparable to wild-type Rev, consistent with the CDMS data (Fig. 3F). These observations strongly indicate that Rev is able to tolerate greater variability to mutation of Crm1 binding residues than expected and that the loss of activity observed in our reporter assay is still above some functional minimum, or that other mechanisms, present in the context of viral replication, compensate for this loss of activity.

The OD, ARM and NES are all structured, critical regions of Rev, where a point mutation is often sufficient to abolish function. While our CDMS data indicate that many residues of the OD and NES are tolerant to

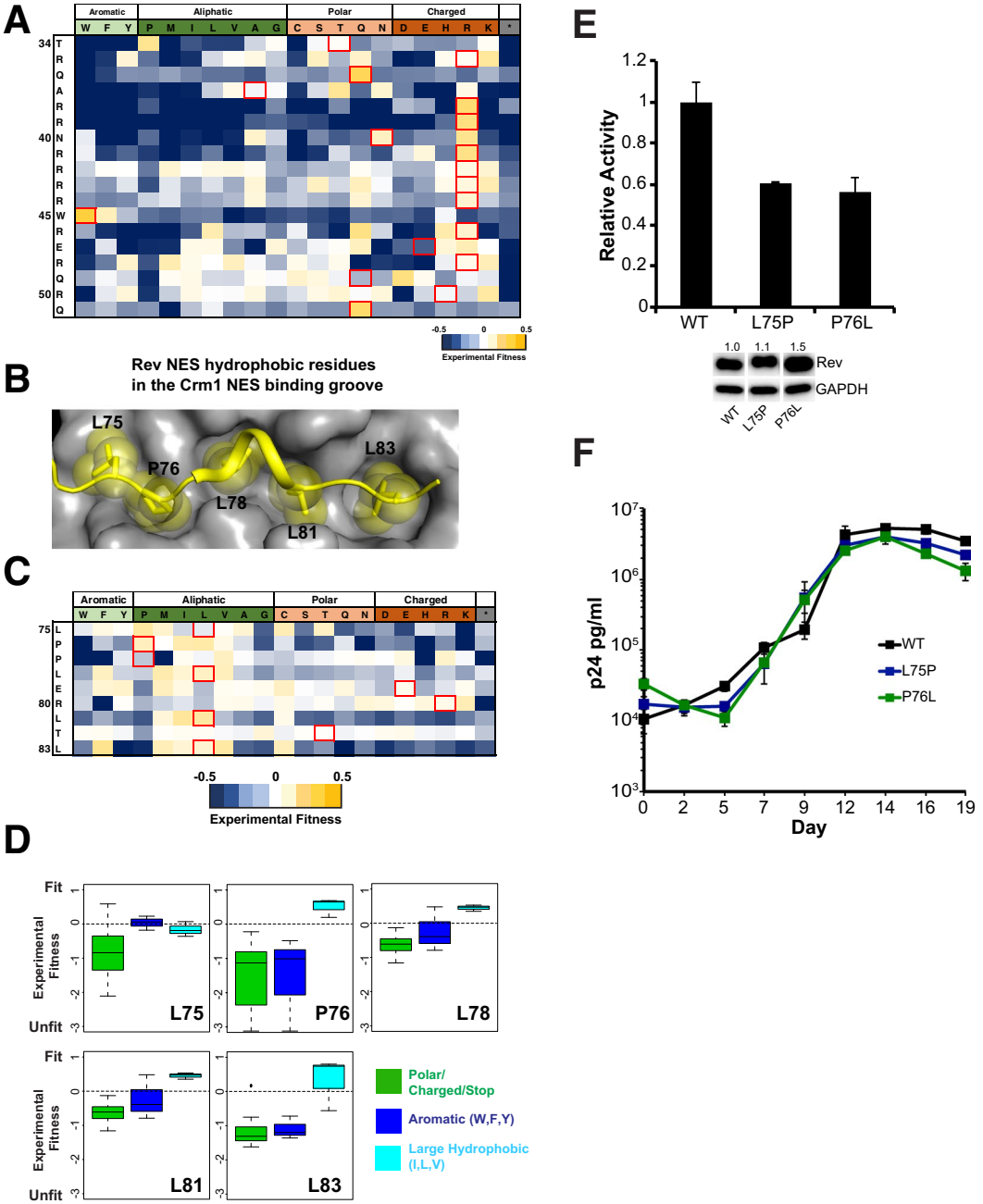

Figure 3. Mutational limits of the Rev ARM and NES. (**A**) CDMS fitness values for residues in the Rev ARM[10] (**B**) Structure of Rev NES (yellow) bound to Crm1 (grey) (PDBID: 3NBZ) with the Rev hydrophobic residues binding to the Crm1 NES binding groove shown as spheres (**C**) CDMS fitness values for the NES[10] (**D**) Preferences for Rev NES contact residues using CDMS fitness values[10]. (**E**) Reporter assay probing export activity of Rev NES mutations. Data are mean ± standard deviation (s.d) of biological replicates. Western blots below show protein expression for Strep-tagged Rev and GAPDH as loading control, and numbers above blots represent quantified band intensities for Rev mutants relative to WT-Rev. The original blot for the cropped bands is shown in Fig. S5. (**F**) Viral replication profiles of viruses containing Rev with NES mutations. Data are mean ± s.d of biological triplicates.

substitutions, conferring mutational robustness/plasticity, the ARM has more stringent sequence preferences. This may reflect the more complex nature of interactions mediated by the ARM (polar, charged, directional as well as RNA and protein), making it relatively more fragile than the OD or NES.

The N- and C-termini influence Rev stability. While the function of the ARM, OD and NES are fairly well understood, the roles of the linkers in Rev function, apart from connecting the modules, is less clear. Since most residues within these regions do not experience strong selection for a specific amino acid/category, point

mutational datasets are insufficient to determine their sequence constraints. Thus, we constructed a series of truncation and gross substitution mutations within these regions and measured their activities to interrogate their functions more rigorously.

The N-terminal 8 residues of Rev are disordered in most crystal structures (Fig. 1B)[13–15,29], but a recent structure of a scFv-bound Rev solved at 2.3 Å shows that the first helix begins at residue 5 (PDB ID: 5DHV)[29]. Genetic screens and selection assays have identified residue 12 as the first essential contact in the Rev OD[27,28]. The presence of an overlapping reading frame encoding the critical ARM of Tat, another essential viral protein, has confounded conservation analyses of the Rev N-terminus. Indeed, previous work suggests that this region does not experience strong selective pressure (excluding requirements for translation initiation), but is well conserved in Rev in order to maintain Tat function[10].

To more fully investigate the role of N-terminus of Rev, we measured the effects of truncations in our export assay. Deleting the first 10 residues (Rev Δ1–10) had no effect on activity whereas any further truncation resulted in loss of activity (Fig. 4A). Replication spreads of viruses containing these truncations recapitulated the export results, with the virus expressing Rev Δ1–10 replicating similar to the wild-type Rev virus, whereas the Rev Δ1–11 virus failed to replicate (Fig. 4B). These results corroborate that the first 10 residues of Rev are not required for function.

Unexpectedly, none of the truncated Rev proteins could be detected by western blot, despite some (such as Rev Δ1–10) being functional in both reporter and replication assays (Fig. 4A,B). Previous reports have shown that Rev function is robust even at very low expression levels[36,37]. Indeed, we observed that serial dilutions of Rev plasmid demonstrated high reporter activity at even very low levels and that activity was actually reduced at higher plasmid levels (Fig. S2A). Therefore, we inferred that the truncated Rev proteins were being expressed at levels sufficient for function but below the level of detection. There was no significant change in the stability of truncated-Rev expressing mRNA (Fig. S2B) indicating that the truncation affects only protein stability. We were able to detect low levels of truncated Rev in the presence of proteasome inhibitor, MG132 (Fig. 4C), suggesting that the truncated proteins were being degraded by the proteasome. We next made smaller truncations to determine at which point Rev expression was restored and found that Rev Δ1–8 showed modest expression whereas Rev Δ1–5 showed near-wild-type expression. The results suggest that the truncations affect the ability of the first helix to form completely and stably (Fig. 4D), thereby influencing protein stability. Remarkably, early mutational work on Rev also identified point mutations in the N-terminus (R4S5 to D4L5) that reduced Rev stability without completely abolishing function[38].

We also monitored Rev expression from NL4–3 viruses in infected T-cells, a more relevant condition. We utilized Rev with a C-terminal Flag tag (for ease of detection) in *rev-in-nef* viruses[15], infected SupT1 cells with these viruses and tracked Rev expression in cell lysates. Rev Δ1–8 displayed poorer expression than full-length Rev, confirming the reduced stability of Rev in T-cells (Fig. 4E), yet conferred no defect in replication. Overall, our findings indicate that the N-terminal residues are not required for replication but nevertheless contribute to Rev stability.

The Rev C-terminus (residues 70–116) is disordered in crystal structures although recent work has suggested that it may become ordered in certain contexts. Cryo-EM studies have suggested that residues 94–116 become ordered when Rev forms filaments[29], and C-terminal truncations (1–69 and 1–93) do not form filaments under the same conditions. Evolutionary coupling analysis also suggests that the C-terminus can engage in long-range contacts and become structured[39]. Early truncation studies on Rev at various positions beyond the nuclear export sequence indicated that truncated Rev was functional[38,40–42]. CDMS data show that selection against stop codons ends at residue 86 (the NES ends at residue 83) (Fig. 5A) indicating that the C-terminus is not required for viral replication. Curiously, the distribution of fitness values for stop codons between position 86–116 is overall positive (Fig. 5B; red distribution is right shifted), indicating that viruses with a truncated Rev might have a competitive advantage compared to wild-type virus and suggesting an inhibitory role for the Rev C-terminus[39]. This distribution is distinct from that of the C-terminus of Tat, another overlapped essential HIV protein, where the fitness values of stop codons appear to be near neutral after the last functional Tat domain at residue 65 (Fig. 5B; blue distribution centered around zero).

C-terminal truncations (Rev 1–85, Rev 1–83) are functional in our reporter and replication assays, further confirming that this region is dispensable for function, while truncation of a single residue in the NES (Rev 1–82) results in a dramatic loss of function (Fig. 5C,D). Similar to the N-terminal truncations, we were unable to detect Rev protein with C-terminal truncations despite functional reporter activity, with no significant change in the stability of truncated-Rev mRNA (Fig. S2B), and observed low levels of truncated Rev in the presence of proteasome inhibitor, MG132 (Fig. 5E). However, unlike the N-terminus, smaller truncations did not restore Rev expression as 1–85, 1–90 and 1–100 all express poorly (Fig. 5F). Intriguingly, previous studies of C-terminal truncations in COS cells did not result in reduced protein stability[38]. Rev 1–85 also displayed reduced stability in virus-infected T-cells, although the virus replicates as well as the wild-type Rev virus (Fig. 5G). Thus, the C-terminus, like the N-terminus, is not required for RNA export or viral replication but regulates protein stability in a manner whose mechanism and functional consequence are not yet understood.

Substitutions to the Rev turn produce different effects on RNA export and viral replication. Rev contains an 8-residue proline rich turn which separates the first two helices and spans the exon junction in the viral genome (Fig. 1A,C). Recent structural studies implicate the turn residues as part of an additional Rev-Rev interaction surface where residues W45, P28, P29 and P31 "hook" with the same residues on an adjacent Rev molecule. This interface has been proposed to act as a molecular bridge between independent oligomerization events along the RRE[29,30]. Most residues in the turn do not have strong preferences in the CDMS study (Fig. 1B)[10], consistent with our understanding of the region as a simple linker. However, residue 28 demonstrated a strong preference for an aromatic side chain (W,F,Y) rather than the reference NL4–3 proline (P) (Fig. S3).

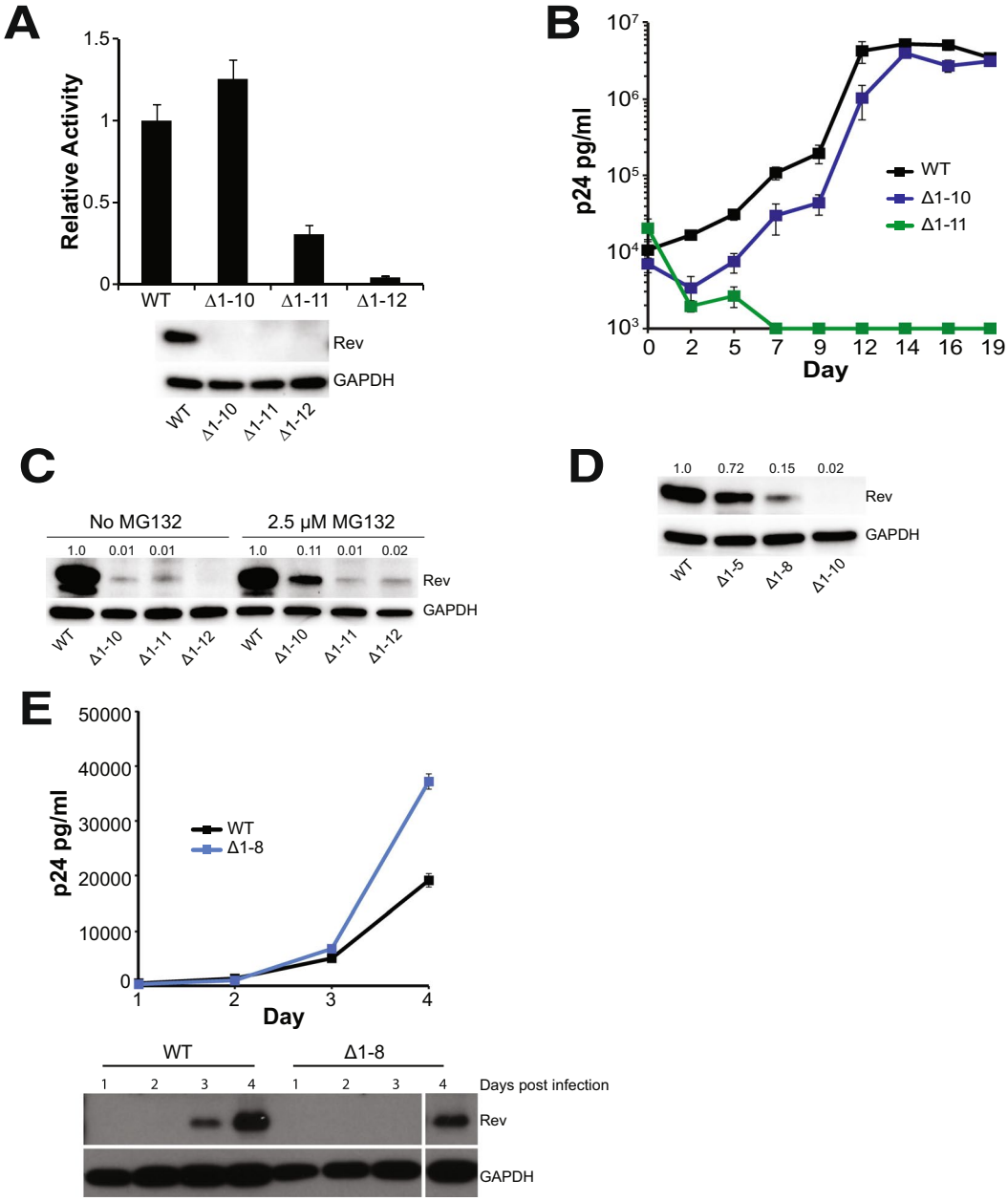

Figure 4. The HIV-1 Rev N-terminus is not required for function but regulates Rev stability. (**A**) Reporter assay probing export activity of Rev N-terminal truncations. Data are mean ± standard deviation (s.d) of biological replicates. Western blots below show protein expression for Strep-tagged Rev and GAPDH as loading control. (**B**) Viral replication profiles of viruses containing Rev with N-terminal truncations. Data are mean ± s.d of biological triplicates. Viral p24 levels below 1000 pg/ml are shown as 1000 pg/ml in the plots for illustration purposes. (**C**) Adding MG132, a proteasome inhibitor partially restores expression of N-terminally truncated Rev in transiently transfected 293 T cells. Numbers above blots represent quantified band intensities for Rev mutants relative to WT-Rev. (**D**) Shortening the length of N-terminal truncations restores protein expression transiently transfected 293 T cells. Numbers above blots represent quantified band intensities for Rev mutants relative to WT-Rev. (**E**) N-terminal truncations affect Rev expression/stability from HIV-1 (with Flag-tagged Rev in the *nef* locus) in infected T-cells. Plot shows virus production from cell supernatant, quantified by p24 ELISA. Western blots below show protein expression for Flag-tagged Rev and GAPDH loading control from the cells. The original blot for the cropped bands is shown in Fig. S5. Data are mean ± s.d of biological replicates.

Interestingly, the most common allele in patient sequences is Y28, although P28 is also substantially represented and is found in many lab-adapted HIV-1 strains[43]. It is possible that a P28Y substitution could provide stacking interactions to stabilize this aromatic interface within the turn[30], but we found no significant difference in activity between the proline and tyrosine variants (Fig. S3). Even an alanine substitution at position 28 is only marginally

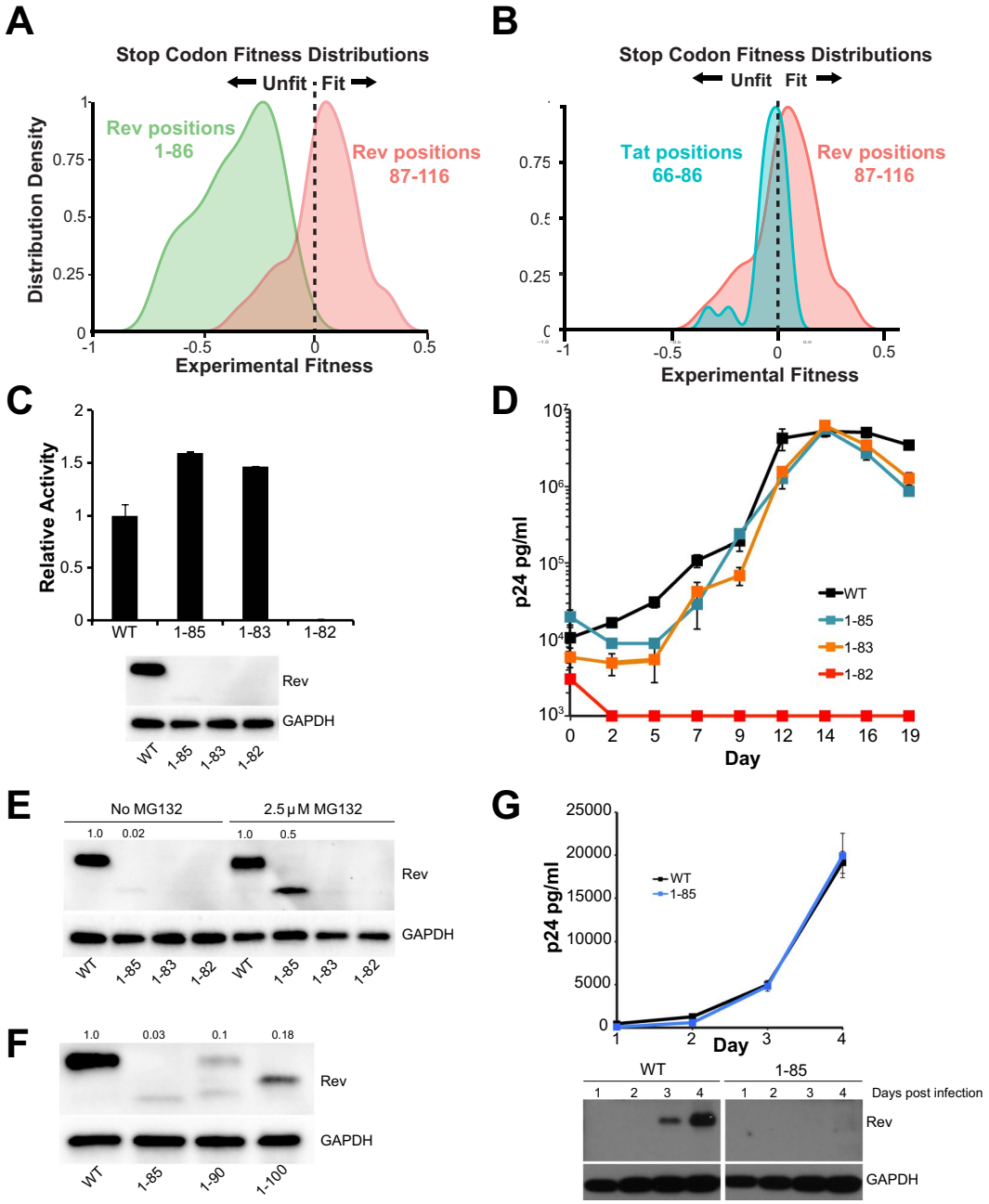

Figure 5. The HIV-1 Rev C-terminus is not required for function but regulates Rev stability. (**A**) Stop codon fitness distributions from CDMS dataset for Rev[10]. (**B**) Stop codon fitness distributions from CDMS dataset for Tat C-terminus (blue, residues 66–86) and Rev C-terminus (red, residues 86–116) (stop codon data for both proteins[10]. (**C**) Reporter assay probing export activity of Rev C-terminal truncations. Data are mean ± standard deviation (s.d) of biological replicates. Western blots below show protein expression for Strep-tagged Rev and GAPDH as loading control. (**D**) Corresponding viral replication spread experiments. Viral p24 levels below 1000 pg/ml are shown as 1000 pg/ml in the plots for illustration purposes. Data are mean ± s.d of biological triplicates. (**E**) Adding MG132, a proteasome inhibitor partially restores expression of Rev with C-terminal truncations in transiently transfected 293 T cells. Numbers above blots represent quantified band intensities for Rev mutants relative to WT-Rev. (**F**) Shortening the length of C-terminal truncations does not restore protein expression transiently transfected 293 T cells. Here, we note that we consistently observe two bands for Rev 1–90, one with a faster mobility and the second with the same mobility as full-length Rev. Numbers above blots represent quantified band intensities for Rev mutants relative to WT-Rev. (**G**) C-terminal truncations affect Rev expression/stability from HIV-1 (with Flag-tagged Rev in the *nef* locus) in infected T-cells. Plot shows virus production from cell supernatant, quantified by p24 ELISA. Western blots below show protein expression for Flag-tagged Rev and GAPDH loading control from the cells. The original blot for the cropped bands is shown in Fig. S5. Data are mean ± s.d of biological replicates.

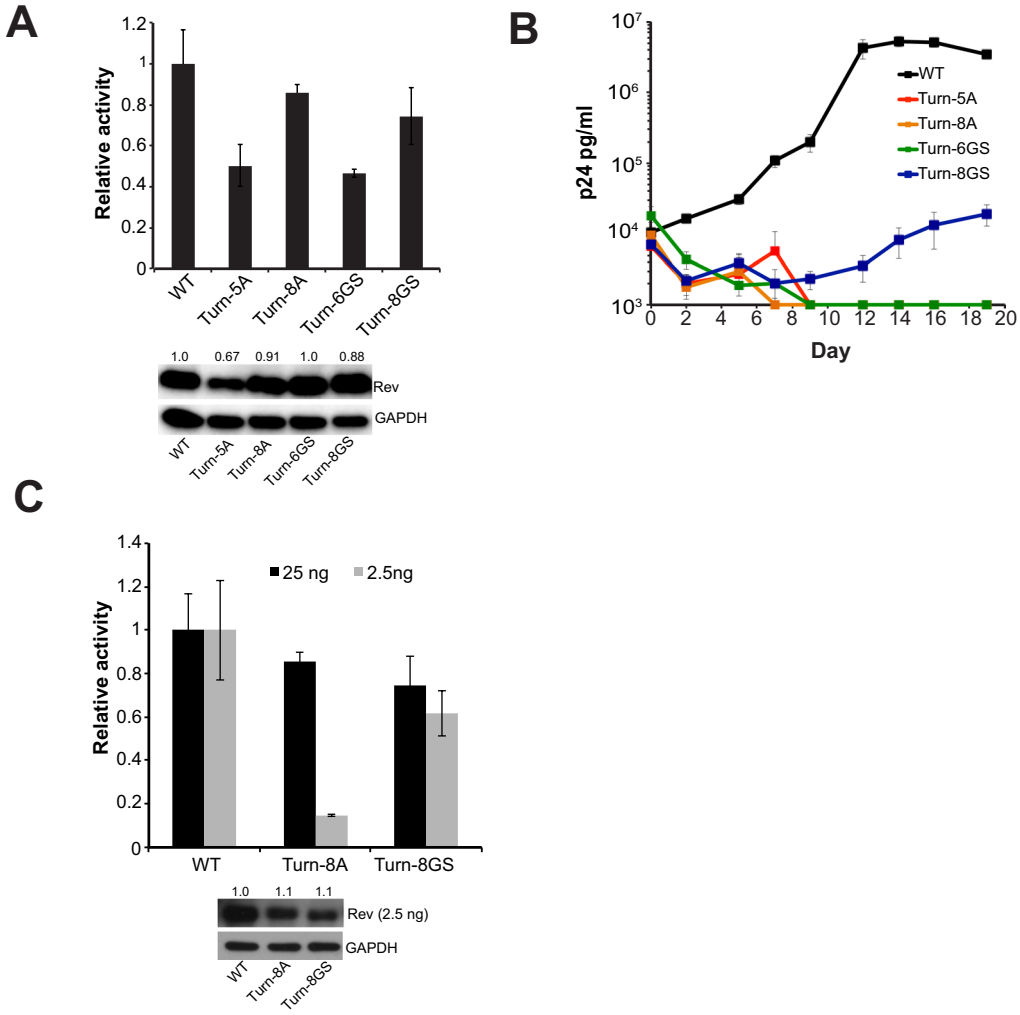

Figure 6. The turn plays an intriguing role in Rev function (**A**) Reporter assay monitoring function of Rev constructs with mass substitution of turn residues 26–33. Data are mean ± standard deviation (s.d) of biological replicates. Western blots below show protein expression for Strep-tagged Rev and GAPDH as loading control. Numbers above blots represent quantified band intensities for Rev mutants relative to WT-Rev. The Turn-5A mutant shows reduced Rev expression suggesting that it might be less stable than the other mutants. (**B**) Corresponding viral replication spread experiments. Viral p24 levels below 1000 pg/ml are shown as 1000 pg/ml in the plots for illustration purposes. Data are mean ± s.d of biological triplicates. (**C**) Reporter assay monitoring export activity at 25 ng and 2.5 ng of transfected Rev plasmid. Western blots below showing protein expression for Strep-tagged Rev (at 2.5 ng only) and GAPDH as loading control. Data are mean ± standard deviation (s.d) of biological replicates and numbers above blots represent quantified band intensities for Rev mutants relative to WT-Rev. Export activity for both quantities of transfected Rev before normalization are shown in Fig. S3.

weaker in replication assays, indicating substantial sequence plasticity, but leaving the manner in which Y28 confers a fitness advantage in competition experiments and in patients still unclear. Nevertheless, given that the turn appeared mutationally plastic in our CDMS studies and since it also determines the relative orientation of the first two helices, we examined its mutational limits with respect to side chain preference, length and structure.

To this end, we created a set of mutations that replaced the turn residues (26–33) with either all alanines (turn-8A, designed to promote helical propensity[44]) or glycines and serines (turn-8GS, designed to allow greater flexibility), and monitored the effects on Rev function in both reporter and replication assays. Surprisingly, we found that both variants yielded significant levels of export in the reporter assay (Fig. 6A) although neither was as active as wild-type Rev. However, shortening the length of the substitutions reduced export activity indicating that the turn must be of sufficient length to connect the two helices and produce functional Rev.

Despite significant activities in the reporter assay, all turn mutants were defective in replication, with only the turn-8GS mutant-virus showing any replication although severely attenuated (Fig. 6B). Given our previous observation that Rev is functional at even low concentrations, we repeated the reporter assay with 10-fold less Rev plasmid (2.5 ng versus 25 ng) to determine if less Rev expression might recapitulate the replication results. Indeed, the export activity of the turn-8A mutant dropped significantly at low Rev expression levels (Fig. 6C), indicating that it is defective and probably structurally unsuitable due to the replacement of a turn with a helical

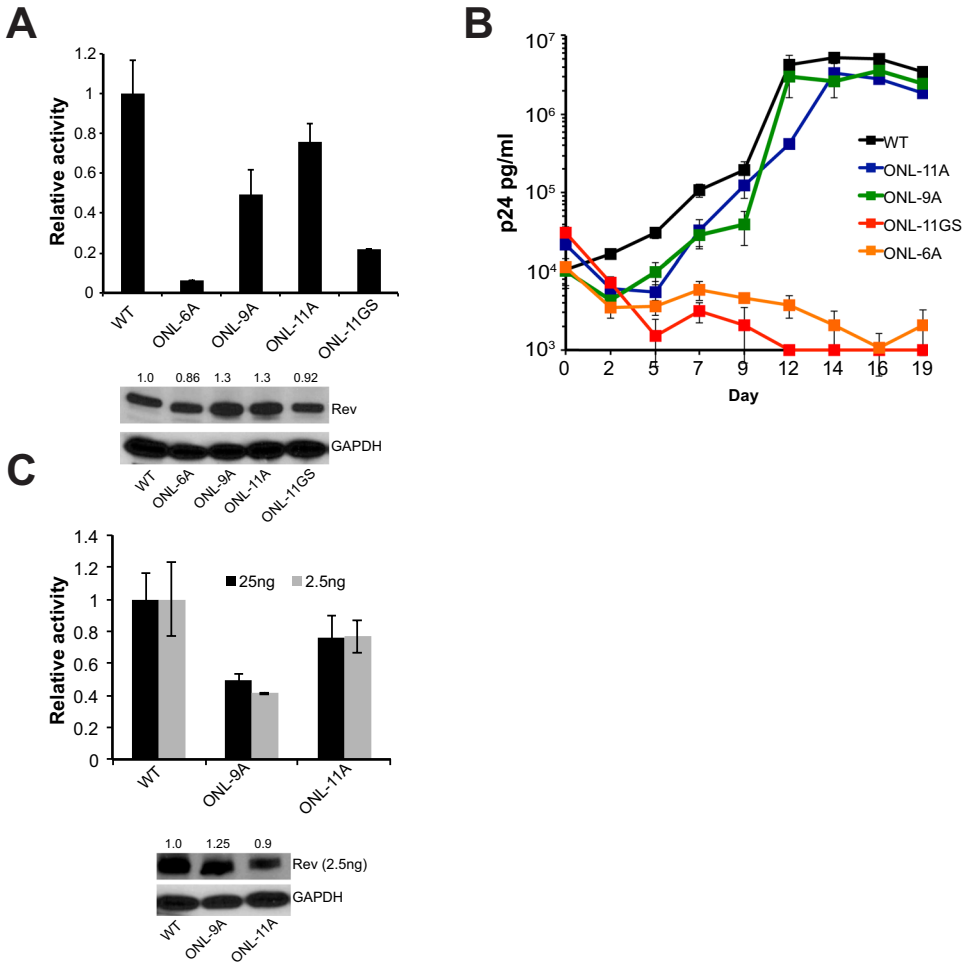

Figure 7. The OD-NES linker (ONL) needs to be structured for Rev function. (**A**) Reporter assay monitoring export function of Rev constructs with mass substitution of ONL residues 62–72. Data are mean ± standard deviation (s.d) of biological replicates. Western blots below showing protein expression for Strep-tagged Rev and GAPDH loading control and numbers above blots represent quantified band intensities for Rev mutants relative to WT-Rev. (**B**) Corresponding viral replication spread experiments. Viral p24 levels below 1000 pg/ml are shown as 1000 pg/ml in the plots for illustration purposes. Data are mean ± s.d of biological triplicates. (**C**) Reporter assay monitoring export activity at 25 ng and 2.5 ng of transfected Rev plasmid. Western blots below showing protein expression for Strep-tagged Rev (at 2.5 ng only) and GAPDH as loading control. Data are mean ± standard deviation (s.d) of biological replicates. Numbers above blots represent quantified band intensities for Rev mutants relative to WT-Rev. Export activity for both quantities of transfected Rev before normalization are shown in Fig. S3.

region. These results are consistent with previous reporter assays showing loss of function when 27PPP29 was mutated to alanines[45]. However, the turn-8GS mutant remained active at the lower Rev plasmid level unlike the turn-8A mutant (Fig. 6B), despite wild-type Rev being 2-fold more active at the lower plasmid level (Fig. S2). This indicates that the turn need only provide flexibility between Rev helices in order for proper export although its export competence is not sufficient for viral replication.

The Rev OD-NES Linker (ONL) becomes structured. An ~11 residue linker region (ONL) connects the second helix of Rev and the NES and is disordered in most Rev crystal structures[14,29]. CDMS data suggests that the residues in this region can sample many amino acids with no clear preference for chemical property (Fig. 1C). In order to probe the requirements of length and structure for this region, we applied a strategy similar to that used to probe the turn. We replaced residues 62–72 of the ONL with either 11 alanines (ONL-11A, to mimic a rigid, structured linker) or 11 glycine-serines (ONL-11GS, to mimic a flexible disordered linker) and tested the effect of these substitutions in reporter and replication assays. ONL-11A was nearly fully functional in the reporter assay whereas ONL-11GS showed a marked decrease in activity (Fig. 7A). Viral replication spreads recapitulated this reporter data (Fig. 7B), with only the structured alanine linker displaying activity. In order to test the length requirements of this region, we shortened the ONL to nine (ONL-9A) or six (ONL-6A) alanines. ONL-9A displayed reduced function in the reporter assay but was surprisingly functional for replication, whereas ONL-6A was poor in both assays (Fig. 7A,B), demonstrating that the ONL has a minimum length requirement.

The activity levels of the ONL-9A and ONL-11A at reduced Rev expression levels (2.5 ng of Rev plasmid) were comparable to those obtained at 25 ng of Rev plasmid, demonstrating that this behavior holds for a broad range of Rev expression levels (Fig. 7C).

The ability of the ONL to tolerate complete replacement with alanines but not glycine/serine suggests that this region becomes structured during Rev function, possibly during oligomeric assembly on the RRE or Crm1 binding. This finding is the first evidence that this region of Rev becomes structured during export; this behavior likely went undetected previously as prior approaches, such as our point mutational CDMS, typically examined the effect of individual point mutations, which alone appear insufficient to disrupt the induced structure.

Discussion

HIV Rev has been extensively investigated over three decades. Structural studies over the past several years have led to a model in which Rev is composed of a few highly conserved, structured regions with well-defined functions that are connected by disordered linkers (Fig. 1A). The results presented here expand the functional repertoire to these linker regions. Indeed, recent work in other systems has demonstrated the importance of disordered and linker regions for establishing and tuning protein interactions[46–48], and particularly for modulating RNA-protein interactions[49].

The use of disordered linkers to tune Rev function may reflect some of the selective pressures imposed by having its entire coding sequence overlapped with other HIV-1 genes, specifically *tat* and *env* (Fig. 1A). In previous work, we observed a complete segregation of function within the overlapped region of *tat* and *rev*, where the overlapped nucleotides encoded amino acids critical for the function of one protein or the other, but never both. Furthermore, the genetically fragile activation domain of Tat is encoded in a non-overlapped portion of the *tat* gene, consistent with the idea that brittle domains in which many mutations are expected to be lethal are unlikely to evolve within regions that must simultaneously maintain two overlapping coding sequences. Since Rev overlaps with other genes over its entire sequence, it apparently evolved relatively mutable interaction surfaces, such as the OD, to create genetic stability. The linker regions modulate Rev function and stability yet are even more mutable and as a result are effectively segregated at the coding level, imposing few constraints on the other reading frames.

Truncations of the N- and C-termini revealed that these regions greatly influence Rev protein stability, yet have little effect on Rev function in the context of our cellular reporter and replication assays. Proteomics studies have reported similar protein interaction profiles for N-terminal Rev truncations (Δ1–8) and wild-type Rev[25], indicating that N-terminus likely does not serve as a major interaction scaffold. However the N-terminus can be phosphorylated[50], and may have regulatory effects that effect the function and/or stability of the protein in patient viral replication settings.

Intriguingly, the CDMS data suggest a fitness advantage to C-terminal truncations and thus a possible inhibitory role (Fig. 5A,B). A recent study proposed that the Rev NES can be masked in the cytoplasm, controlling Rev trafficking in and out of the nucleus[36]. We speculate that removing C-terminal residues might decrease the masking and thereby promote export of viral RNAs to provide a fitness advantage. Conversely, we observe that truncation of the C-terminus decreases protein stability, which might seem to negate any gains to export activity, however very little Rev is required for function, excessive amounts of Rev can lead to reduced export activity (Fig. S2), and viruses harboring undetectable amounts of Rev can replicate well (Figs 4E, 5G). Hence, it appears that truncating the C-terminus might increase fitness by reducing Rev levels to a more optimal level. However, we note that our viral replication experiments are performed with *rev-in-nef*, and *nef* transcripts are nearly two-fold more abundant than *rev* transcripts[51,52]. Therefore, some of the mutants with unaffected replication but reduced Rev stabilities might be compensated for via increased expression of Rev from the *nef* locus. Expression of truncated Rev from its endogenous locus might reveal the true limits for Rev expression and viral replication. However, the Rev C-terminus entirely overlaps with Env (Fig. 1A), and selection pressure from the Env reading frame (either functional requirements or immune targeting of specific Env epitopes) can influence Rev sequence and make stop codons unfavorable at most positions.

The Rev C-terminus length varies considerably amongst HIV-1 isolates (Fig. S4). Apart from the most common 116-amino acid form of Rev, some are 123 amino acids long due to a QSQGTET insertion between residues 95 and 96; other isolates carry this insertion as well as a premature stop codon and produce a truncated, 107-amino acid Rev. Insertion of the QSQGTET sequence reduces export activity (Fig. S4), further suggesting that a long C-terminus inhibits spread, although further studies will be needed to reveal the exact mechanism and functional consequences of Rev C-terminus length.

Given Rev's role in regulating HIV kinetics, multiple mechanisms might fine-tune Rev function within the context of replication. Indeed, the logic of gene expression networks involving Rev continues to be of major interest to the HIV community, and has major implications for the mechanisms of HIV latency[53–55]. Our findings on Rev stability in a cell culture system therefore provide important parameters and considerations for future studies dedicated to deciphering viral kinetics and decoupling the selective pressures of Rev function from other selective pressures of viral replication in more complex systems.

The turn between the helices of Rev is highly mutable but does have certain requirements of length and flexibility. The ability of the turn-8GS mutant to export well in the reporter assay yet fail to replicate suggests that this region may perform additional roles in viral replication. Alternately, the different results in different contexts could also be because of the cell types (and possibly, host factors for Rev interaction) used for reporter assays (293 T cells, transformed kidney cells) and viral replication assays (SupT1 cells, lymphoblastic T-cells). In any case, these results suggest that the turn may form an important functional interface. Recent structural studies have determined that the turn can form a homotypic interaction interface[29,30], and mutations expected to stabilize this interface (P28Y) are enriched in both our CDMS data and patient data[10]. Further characterization will be required to determine the exact role of this interface and more generally the turn itself, in Rev function.

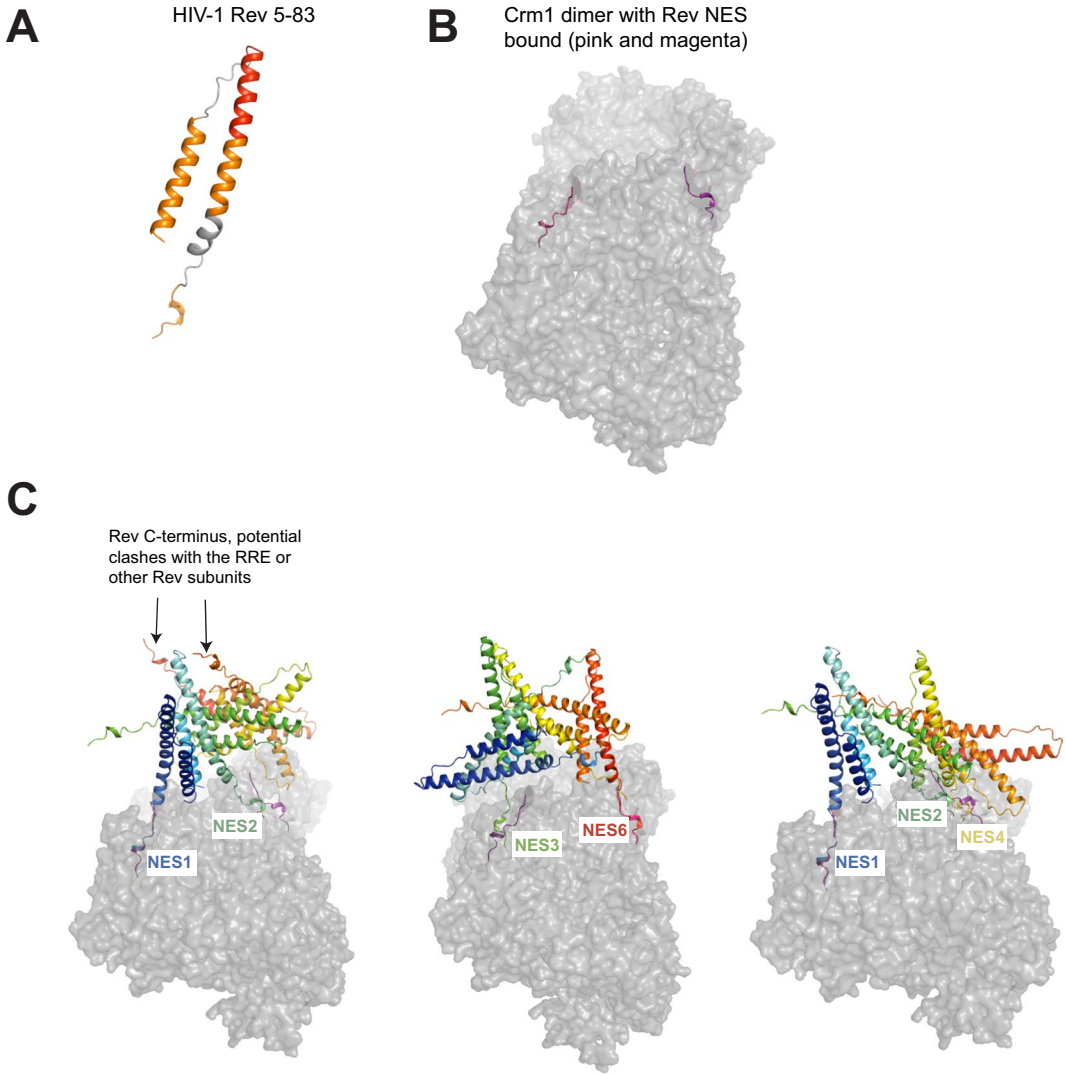

Figure 8. Models of HIV-1 Rev 5–83 and potential interactions of Rev hexamers with a Crm1 dimer. (**A**) Model of HIV-1 Rev 5–83 built using Chimera. (**B**) Model of a Crm1 dimer bound to Rev NES (from[33]) (**C**) Three representative Rev hexamers generated using Rev 5–83 and dimerization and oligomerization surfaces (as in[15]). Rev subunits are colored from blue (subunit 1) to red (subunit 6). Different Rev NES sequences were aligned with one of the Crm1-bound Rev-NES peptide: NES1 from the first panel, NES6 from the second and NES 1 from the third panel. We then observed which other Rev NES sequences were in close proximity to the second Crm1-bound Rev-NES peptide.

Similarly, our manipulation of the ONL uncovered that it does not tolerate replacements that increase its propensity for disorder, suggesting that it becomes structured during some step of export complex formation. Since this stretch of residues can be replaced with alanines without significant loss of function, it appears that these residues do not form a specific interaction surface, but instead play a structural role, in which the specific sequence itself is unimportant as long as the region can adopt a presumably helical structure. Negative-stain electron microscopy has shown that a Crm1 dimer binds to the Rev-RRE complex with the two NES-binding grooves distinctly positioned to associate with two NESs from the complex[33]. A structured ONL might reduce the entropic penalty associated with Crm1 assembly and also fix NESs into a favorable orientation for this interaction, thereby improving the energetics of export complex formation.

To examine how a structured ONL might impact Rev-RRE-Crm1 complex formation, we re-visited previous models for Rev-RRE-Crm1 complex assembly[15]. Given the established plasticity in RRE structure and the hydrophobic Rev interaction surfaces, we proposed earlier that these features allow Rev-RRE complexes with varying sequences to adopt configurations favorable for Crm1 recruitment and nuclear export[15]. Adding the requirement that the OD-NES linker needs to be structured (helical with the limitations imposed by two prolines in the NL4–3 sequence), we built a model of a Rev monomer containing residues 5–83 (Fig. 8A), where 83 is the last residue of the NES, from the structures of Rev (5DHV)[30] and a Rev NES-Crm1 complex (3NBZ)[17] using Chimera[56]. Previous studies established a hexamer of Rev on a minimal, functional 240 nucleotide RRE[19]. Therefore, we built three representative Rev hexamers using Rev 5–83 as described previously[15] and aligned the NES from one Rev

subunit in the hexamer to a Rev-NES peptide bound to Crm1 in a model of Crm1-dimer-Rev-NES complex[33] (Fig. 8B,C). Four main aspects become apparent from these models: (1) With more structured residues of Rev past its second helix, only a subset of hexamers are plausible due to steric clashes of the Rev C-terminus with other Rev subunits, the RRE or Crm1 (Fig. 8C, left panel); (2) The RNA-bound Rev dimer, with its narrow crossing angle of 50° between subunits, places the two NESs in a nearly favorable configuration to associate with a Crm1 dimer (Fig. 8C, left and right panels); (3) Different hexamer configurations can recruit Crm1 dimers using different combinations of Rev subunits (Fig. 8C); and (4) In some hexamer configurations, more than one NES is positioned close to an NES-binding groove in one of the Crm1 subunits (Fig. 8C, right panel). Given that the RRE enhances Rev binding to Crm1[33], these configurations may in fact be the dominant class of hexamers, forming an ensemble of nuclear export complexes wherein multiple NESs may associate with Crm1 in a variety of orientations[57–59]. In this way, the structuring of the ONL may restrict the possible orientations of Rev-RRE hexamers to a subset of export-competent configurations.

By our systematic analysis of each region of Rev, we conclude that every region of the protein influences its function, structure, or stability, consistent with optimal usage of limited genetic information. In the context of the viral genome and its overlapped reading frames, the virus is able to segregate essential functional and structural motifs and yet further tune Rev function using linker regions with relatively relaxed sequence and length requirements[10]. Indeed, recent analyses of Rev-like proteins in other retroviruses suggest that many key structural features - such as the key hydrophobic residues in the OD – remain conserved even as the primary sequence diverged[60]. Thus, the nuclear export function of Rev can be achieved through a wide variety of sequence configurations. Furthermore, our experiments, as well as studies of patient isolates[61], have demonstrated a wide range of Rev activity and expression levels, providing evidence of an additional levels of robustness beyond the mutational plasticity that governs protein stability and interactions studied in this work.

This study and others[62–64] show that reporter and replication assays largely correlate, suggesting that RNA export activity is the major role of Rev in the viral life cycle; however it is interesting that some Rev mutants exhibit different behaviors. For example, Rev L75P, P76L, and ONL-9A all have lowered reporter activities (40–60% of wild-type Rev) but are completely replication competent. Conversely, turn-8GS has export phenotype comparable to the above constructs but is completely defective for replication. These results illustrate that not all export defects lead to viral replication defects, suggesting that regulatory mechanisms may exist to compensate for these defects, and that reporter export activity does not always fully explain the role of Rev in the viral life cycle, raising the possibility that Rev may possess still unknown functions. Other roles for Rev, such as regulating splicing[65], translation[66–68] and encapsidation[69–71] of RRE-containing messages have been reported as well as other interactions with host partners[26,72–75], and many of the mutants described here may help elucidate those roles.

Future work in the endogenous HIV-1 genomic context, as well as a variety of infectable cell types may shed light on these questions and reveal additional evolutionary constraints that shape Rev's robustness. Such work will require careful design as the kinetics of Rev expression and viral spread within an organism are complex, and are regulated on multiple levels. In many of these systems, selective forces do not necessarily act to produce only a high viral titer as they do in cell culture models. The work presented here should prove valuable in deconvoluting the functional consequences of mutations in these more complex systems. In sum, our results add to an increasingly mature structural picture of Rev macromolecular complexes and demonstrate that large-scale mutagenesis approaches combined with targeted assays and structural knowledge can reveal novel and subtle mutational constraints in even well-studied systems.

Materials and Methods

Construction of mutants. Primers harboring the appropriate mutations were chemically synthesized (IDT) and then used in a 2-step PCR reaction and inserted into pcDNA4TO-2xStrep-Rev (Invitrogen mammalian expression vector, described in[10]) via Gibson Assembly for testing in export assays. A similar strategy was undertaken to make viral mutants in a *rev-in-nef* virus as described previously[10,15].

Cells and media. HEK293T cells were maintained in DMEM, high glucose (with 10% heat-inactivated fetal bovine serum (FBS) and 1% penicillin/streptomycin) at 37°C and 5% CO_2. SupT1 cells were maintained in RPMI (with 25 mM HEPES pH 7.4, 10% heat-inactivated fetal bovine serum (FBS) and 1% penicillin/streptomycin) at 37°C and 5% CO_2.

Rev reporter assays. Reporter assays to monitor Rev export activity were carried out in HEK 293 T cells (from ATCC) by transient transfection. In a typical experiment, 293 T cells in a 24-well plate were transfected with 250 ng of a pCMV-GagPol-RRE reporter construct[35], 25 ng of pcDNA4TO-2xStrep-Rev (Invitrogen mammalian expression vector) (or indicated amount) and 10 ng pcDNA4TO-Firefly luciferase (to monitor transfection efficiency) using Polyjet transfection reagent (SignaGen). We chose these quantities for Rev plasmid to maintain accuracy in plasmid concentrations across different constructs and for easily detectable protein expression. Forty-eight hours after transfection, the culture supernatant was removed, the cells were lysed in Tris-buffered saline containing 1% Triton X-100 and protease inhibitors, and intracellular Gag (p24) was quantified by ELISA. Rev expression levels were verified by western blot using HRP-conjugated-anti-StrepTagII antibody (IBA biosciences) with GAPDH used as a loading control. Blots were detected by chemi-luminescence using ChemiDoc (Biorad); exposure times were adjusted so that the signal was not saturated and bands were quantified using ImageJ[76]. Band intensities were in the linear range for these exposure conditions and for the Rev quantities used in our experiments (Fig. S2). For experiments with MG132, 2.5 uM proteasome inhibitor MG132 was added 30 hours after transfection and cells were harvested 16 hours later.

Measuring cytoplasmic gagpol mRNA levels. In order to confirm that the p24 levels quantified by ELISA report mainly on Rev's nuclear export function and not because of any effects of Rev on Gag translation, we quantified cytoplasmic GagPol mRNA levels in the presence of some of our Rev mutants (Fig. S1). 293 T cells in a 24-well plate were transfected with 250 ng of a pCMV-GagPol-RRE reporter construct, 2.5 ng of pcDNA-4TO-2xStrep-Rev (or indicated Rev mutant) and 10 ng pcDNA4TO-Firefly luciferase using Polyjet transfection reagent (SignaGen). as described above in Rev reporter assays. 24 hours post transfection, the culture supernatant was aspirated, cells were detached using DPBS containing 5 mM EDTA, washed twice in ice-cold DPBS. To lyse the plasma membrane, cells were resuspended gently by flicking in 175 ul of ice-cold buffer RLN (50 mM Tris.Cl pH 8.0, 140 mM NaCl, 1.5 mM MgCl2, 0.5% Nonidet p-40). Following a 5-minute incubation on ice, the lysate was centrifuged at 300xg for 2 minutes at 4°C and the supernatant (cytoplasmic extract) was used for RNA extraction using the Qiagen RNeasy kit. DNase treatment and RT-PCR reactions were set up using the BioRad iScript gDNA Clear cDNA synthesis kit. Relative RNA abundance was quantified by qPCR and the GagPolRRE mRNA levels were normalized to GAPDH mRNA levels (Fig. S1B). Primers used for qPCR were:

GagPolRRE-F: 5′ GGAAACATGGTGGACAGAGTAT 3′
GagPolRRE-R: 5′ CTGCTCCTACTATGGGTTCTTTC 3′
GAPDH-F: 5′ GTCTCCTCTGACTTCAACAGCG 3′
GAPDH-R: 5′ ACCACCCTGTTGCTGTAGCCAA 3′.

Viral replication assays. Rev mutants were cloned into the *nef* region of a NL4–3 proviral construct, mutated to ensure no expression from the endogenous *rev* locus[15]. For generation of viruses, HEK 293 T cells in a 6 well plate were transfected with 1000 ng of NL4–3 proviral plasmid and 20 ng of pCDNA4TO-Rev to promote robust production of viruses containing Rev mutations. Virus supernatants were collected 48 hours after transfection, quantified by p24 capsid ELISA and stored in aliquots at −80 °C. For viral replication assays, one million SupT1 cells were inoculated with 5 ng of p24 virus in 250 ul media containing 1 ug/ml PEI and 8 ug/ml polybrene and centrifuged for 2 hours at 37 °C. Input virus was removed, cells were washed once with 250 ul of PBS, resuspended in 250 ul of media and placed in an incubator at 37 °C, 5% CO_2[15]. 100 ul of virus-containing supernatants were collected every 48–72 hours for 20 days (and equivalent amounts of fresh media were added to cells) for quantification by p24 ELISA. We wish to note that the replication kinetics of an NL4–3 reference virus and an NL4–3 *rev-in-nef* virus are different, with both viruses displaying peak infection at different times (Fig. S1). This is likely due differences in timing and abundance of Rev-encoding messages from an endogenous *rev* locus and a *rev-in-nef* locus.

To monitor expression of Rev-Flag during viral replication, Rev-Flag, Rev Δ1–8-Flag and Rev 1–85-Flag were cloned into the *nef* locus of a NL4–3 proviral construct and viruses were generated as above. Three million SupT1 cells (NIH AIDS Reagent Program) were infected with 150 ng of p24 (viral capsid) in a total volume of 6 ml containing 1 ug/ml PEI and 8 ug/ml polybrene and centrifuged for 2 hours at 37 °C at 1200xg[15]. Input virus was removed, cells were washed with 5 ml of DPBS and resuspended in 6 ml of media (RPMI supplemented with 25 mM HEPES pH 7.4, 10% fetal bovine serum, 1% penicillin-streptomycin. 1.25 ml of cells were withdrawn on days 1–4 post infection, lysed in RIPA buffer (20 mM Tris-Cl pH 6.4, 140 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS), containing protease inhibitors and Flag-tagged Rev was detected by western blot (mouse anti-Flag antibody, Sigma) with GAPDH used as a loading control. Blots were detected by chemi-luminescence by exposure to an X-Ray film (>8 minute long exposures were necessary to detect Rev).

Rev mRNA stability analysis. 293 T cells in a 24-well plate were transfected as described above in Rev reporter assays. 18 hours post transfection, the culture supernatant was aspirated, cells were detached using DPBS containing 5 mM EDTA, washed twice in ice-cold DPBS and the pelleted cell volume (PCV) was measured. The cells were then washed once in 3x PCV of Buffer A (10 mM Tris-HCl, pH 8.0 1.5 mM MgCl2, 10 mM KCl, 0.5 mM DTT, 0.2U/ul Murine RNase inhibitor (New England BioLabs)) and resuspended in 5x PCV of buffer A. After 10 minutes, Triton X-100 was added to the cells to a final concentration of 0.5%, the cells were passed through a 25 G needle and then centrifuged at 4500 g x g, 5 minutes. The supernatant was used for RNA extraction using the Zymo Quick-RNA Mini Prep kit with DNase treatment. RT-PCR reactions were set up using the BioRad iScript Select cDNA synthesis kit using Oligo (dT) primers. RNA abundance was quantified by qPCR and the Rev mRNA levels were normalized to GAPDH mRNA levels. Primers used for qPCR were:

Rev-F: 5′ GACAGGCCCGAAGGAATAG 3′
Rev-R: 5′ CGCAGATCGTCCCAGATAAG 3′
GAPDH-F: 5′ GTCTCCTCTGACTTCAACAGCG 3′
GAPDH-R: 5′ ACCACCCTGTTGCTGTAGCCAA 3′.

CDMS analysis. Fitness values from previous work[10], representing selection strength based on population genetics modeling and changes in allele frequencies during viral competition assays, were used to display the preferences of each amino acid at each position. To determine each position's general hydrophobic preferences, we grouped these fitness values into related bins (e.g. W,F,Y) and plotted their distributions.

Patient isolate analysis. Analysis of patient sequences was performed in a manner as previously described[10]. Specifically, we downloaded high-quality, curated protein alignments from the publicly available Los Alamos HIV Sequence Compendium (http://www.hiv.lanl.gov/) and calculated amino acid conservation for every position in Rev, representing 2,694 Rev sequences. Gaps not found in the reference HXB2 sequence were discounted from the analysis. Patient sequences consist of only viral sequences and no patient-specific information.

Analysis of stop codon fitness. Data from our CDMS experiments[10] were used to generate distributions of Rev residues 1–86 (position 86 was chosen as it is the most C-terminal position which displays selection against the stop codon), Rev residues 87–116, and Tat positions 66–86 (position 66 was chosen as it is the most N-terminal position in Tat that displays neutral or positive fitness for the stop codon). Distributions were plotted in R using ggplot2's density function smoothed with a bandwidth adjustment parameter of 1 (adjust = 1).

Data Availability

Data generated and/or analyzed during the current study are available from the corresponding author on request.

References

1. Wagner, A. Robustness, evolvability, and neutrality. *FEBS Lett.* **579**, 1772–1778 (2005).
2. Elena, S. F. RNA virus genetic robustness: possible causes and some consequences. *Curr. Opin. Virol.* **2**, 525–530 (2012).
3. Rihn, S. J. *et al*. Extreme genetic fragility of the HIV-1 capsid. *PLoS Pathog.* **9**, e1003461 (2013).
4. Xue, B., Mizianty, M. J., Kurgan, L. & Uversky, V. N. Protein intrinsic disorder as a flexible armor and a weapon of HIV-1. *Cell. Mol. life Sci. C.* **69**, 1211–59 (2011).
5. Rausch, J. W. & Le Grice, S. F. J. HIV Rev Assembly on the Rev Response Element (RRE): A Structural Perspective. *Viruses* **7**, 3053–75 (2015).
6. Pollard, V. W. & Malim, M. H. The HIV-1 Rev protein. *Annu. Rev. Microbiol.* **52**, 491–532 (1998).
7. Malim, M. H. *et al*. HIV-1 structural gene expression requires binding of the Rev trans-activator to its RNA target sequence. *Cell* **60**, 675–83 (1990).
8. Cuevas, J. M., Geller, R., Garijo, R., López-Aldeguer, J. & Sanjuán, R. Extremely High Mutation Rate of HIV-1 *In Vivo*. *PLOS Biol.* **13**, e1002251 (2015).
9. Chirico, N., Vianelli, A. & Belshaw, R. Why genes overlap in viruses. *Proc. R. Soc. B Biol. Sci.* **277**, 3809–3817 (2010).
10. Fernandes, J. D. *et al*. Functional Segregation of Overlapping Genes in HIV. *Cell* **167**, 1762–1773.e12 (2016).
11. Mann, D. *et al*. A Molecular Rheostat: Cooperative Rev Binding to Stem I of the Rev-response Element Modulates Human Immunodeficiency Virus Type-1 Late Gene Expression. *J. Mol. Biol.* **241**, 193–2007 (1994).
12. Malim, M. H. & Cullen, B. R. HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: implications for HIV-1 latency. *Cell* **65**, 241–248 (1991).
13. Daugherty, M. D., Liu, B. & Frankel, A. D. Structural basis for cooperative RNA binding and export complex assembly by HIV Rev. *Nat. Struct. Mol. Biol.* **17**, 1337–1342 (2010).
14. DiMattia, M. A. *et al*. Implications of the HIV-1 Rev dimer structure at 3.2 Å resolution for multimeric binding to the Rev response element. *Proc. Natl. Acad. Sci. USA* **107**, 5810–5814 (2010).
15. Jayaraman, B. *et al*. RNA-directed remodeling of the HIV-1 protein Rev orchestrates assembly of the Rev-Rev response element complex. *Elife* **4**, (2014).
16. Dong, X. X. *et al*. Structural basis for leucine-rich nuclear export signal recognition by CRM1. *Nature* **458**, 1136–U71 (2009).
17. Güttler, T. *et al*. NES consensus redefined by structures of PKI-type and Rev-type nuclear export signals bound to CRM1. *Nat. Struct. Mol. Biol.* **17**, 1367–1376 (2010).
18. Monecke, T. *et al*. Structural basis for cooperativity of CRM1 export complex formation. *Proc. Natl. Acad. Sci. USA* **110**, 960–5 (2013).
19. Daugherty, M. D., Booth, D. S., Jayaraman, B., Cheng, Y. & Frankel, A. D. HIV Rev response element (RRE) directs assembly of the Rev homooligomer into discrete asymmetric complexes. *Proc. Natl. Acad. Sci. USA* **107**, 12481–6 (2010).
20. Daugherty, M. D., D'Orso, I. & Frankel, A. D. A solution to limited genomic capacity: using adaptable binding surfaces to assemble the functional HIV Rev oligomer on RNA. *Mol. Cell* **31**, 824–34 (2008).
21. Zapp, M. L., Hope, T. J., Parslow, T. G. & Green, M. R. Oligomerization and RNA binding domains of the type 1 human immunodeficiency virus Rev protein: a dual function for an arginine-rich binding motif. *Proc. Natl. Acad. Sci. USA* **88**, 7734–8 (1991).
22. Kjems, J., Brown, M., Chang, D. D. & Sharp, P. A. Structural analysis of the interaction between the human immunodeficiency virus Rev protein and the Rev response element. *Proc. Natl. Acad. Sci. USA* **88**, 683–7 (1991).
23. Robertson-Anderson, R. M. *et al*. Single-Molecule Studies Reveal that DEAD Box Protein DDX1 Promotes Oligomerization of HIV-1 Rev on the Rev Response Element. *J. Mol. Biol.* **410**, 959–971 (2011).
24. Naji, S. *et al*. Host cell interactome of HIV-1 Rev includes RNA helicases involved in multiple facets of virus production. *Mol. Cell. Proteomics* **11**(M111), 015313 (2012).
25. Hadian, K. *et al*. Identification of a Heterogeneous Nuclear Ribonucleoprotein-recognition Region in the HIV Rev Protein. *J. Biol. Chem.* **284**, 33384–33391 (2009).
26. Hammond, J. A. *et al*. A Survey of DDX21 Activity During Rev/RRE Complex Formation. *J. Mol. Biol.* **430**, 537–553 (2018).
27. Jain, C. & Belasco, J. G. A structural model for the HIV-1 Rev-RRE complex deduced from altered-specificity rev variants isolated by a rapid genetic strategy. *Cell* **87**, 115–25 (1996).
28. Jain, C. & Belasco, J. G. Structural model for the cooperative assembly of HIV-1 Rev multimers on the RRE as deduced from analysis of assembly-defective mutants. *Mol. Cell* **7**, 603–14 (2001).
29. DiMattia, M. A. *et al*. The Structure of HIV-1 Rev Filaments Suggests a Bilateral Model for Rev-RRE Assembly. *Structure* **24**, 1068–1080 (2016).
30. Watts, N. R. *et al*. A new HIV-1 Rev structure optimizes interaction with target RNA (RRE) for nuclear export. *J. Struct. Biol.* https://doi.org/10.1016/J.JSB.2018.03.011 (2018).
31. Henderson, B. R. & Percipalle, P. Interactions between HIV rev and nuclear import and export factors: the rev nuclear localisation signal mediates specific binding to human importin-β. *J. Mol. Biol.* **274**, 693–707 (1997).
32. Battiste, J. L. *et al*. Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science* **273**, 1547–1551 (1996).
33. Booth, D. S., Cheng, Y. & Frankel, A. D. The export receptor Crm1 forms a dimer to promote nuclear export of HIV RNA. *Elife* **4**, (2014).
34. Smith, A. J., Cho, M. I., Hammarskjöld, M. L. & Rekosh, D. Human immunodeficiency virus type 1 Pr55gag and Pr160gag-pol expressed from a simian virus 40 late replacement vector are efficiently processed and assembled into viruslike particles. *J. Virol.* **64**, 2743–2750 (1990).
35. Srinivasakumar, N. *et al*. The effect of viral regulatory protein expression on gene delivery by human immunodeficiency virus type 1 vectors produced in stable packaging cell lines. *J. Virol.* **71**, 5841–8 (1997).
36. Behrens, R. T., Aligeti, M., Pocock, G. M., Higgins, C. A. & Sherer, N. M. Nuclear Export Signal Masking Regulates HIV-1 Rev Trafficking and Viral RNA Nuclear Export. *J. Virol.* **91**, (2017).
37. O'Carroll, I. P. *et al*. Contributions of Individual Domains to Function of the HIV-1 Rev Response Element. *J. Virol.* **91**, e00746–17 (2017).

38. Malim, M. H., Böhnlein, S., Hauber, J. & Cullen, B. R. Functional dissection of the HIV-1 Rev trans-activator–derivation of a trans-dominant repressor of Rev function. *Cell* **58**, 205–214 (1989).
39. Toth-Petroczy, A. *et al*. Structured States of Disordered Proteins from Genomic Sequences. *Cell* **167**, 158–170.e12 (2016).
40. Olsen, H. S., Cochrane, A. W., Dillon, P. J., Nalin, C. M. & Rosen, C. A. Interaction of the human immunodeficiency virus type 1 Rev protein with a structured region in env mRNA is dependent on multimer formation mediated through a basic stretch of amino acids. *Genes Dev.* **4**, 1357–64 (1990).
41. Venkateshand, L. K. & Chinnadurai, G. Mutants in a Conserved Region near the Carboxy-Terminus of HIV-1 Rev Identify Functionally Important Residues and Exhibit a Dominant Negative Phenotype. *Virology* **178**, 327–330 (1990).
42. Malim, M. H., McCarn, D. F., Tiley, L. S. & Cullen, B. R. Mutational definition of the human immunodeficiency virus type 1 Rev activation domain. *J. Virol.* **65**, 4248–54 (1991).
43. Foley, B. *et al*. *HIV Sequence Compendium 2013*. (Theoretical Biology and Biophysics Group 2013).
44. Pace, C. N. & Scholtz, J. M. A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins.
45. Hope, T. J., Huang, X. J., McDonald, D. & Parslow, T. G. Steroid-receptor fusion of the human immunodeficiency virus type 1 Rev transactivator: mapping cryptic functions of the arginine-rich motif. *Proc. Natl. Acad. Sci. USA* **87**, 7787–91 (1990).
46. Borgia, A. *et al*. Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **555**, 61–66 (2018).
47. Hultqvist, G. *et al*. Emergence and evolution of an interaction between intrinsically disordered proteins. *Elife* **6** (2017).
48. Gitlin, L., Hagai, T., LaBarbera, A., Solovey, M. & Andino, R. Rapid Evolution of Virus Sequences in Intrinsically Disordered Protein Regions. *PLoS Pathog.* **10**, e1004529 (2014).
49. Protter, D. S. W. *et al*. Intrinsically Disordered Regions Can Contribute Promiscuous Interactions to RNP Granule Assembly. *Cell Rep.* **22**, 1401–1412 (2018).
50. Meggio, F. *et al*. Phosphorylation of HIV-1 Rev protein: implication of protein kinase CK2 and pro-directed kinases. *Biochem. Biophys. Res. Commun.* **226**, 547–54 (1996).
51. Purcell, D. F. & Martin, M. A. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *J. Virol.* **67**, 6365–78 (1993).
52. Ocwieja, K. E. *et al*. Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic Acids Res.* **40**, 10345–10355 (2012).
53. Felber, B. K., Drysdale, C. M. & Pavlakis, G. N. Feedback regulation of human the Rev protein. Feedback Regulation of Human Immunodeficiency Virus Type 1 Expression by the Rev Protein. **64** (1990).
54. Hansen, M. M. K. *et al*. A Post-Transcriptional Feedback Mechanism for Noise Suppression and Fate Stabilization. *Cell* **173**, 1609–1621.e15 (2018).
55. Likhoshvai, V. A. *et al*. Mathematical model of the Tat-Rev regulation of HIV-1 replication in an activated cell predicts the existence of oscillatory dynamics in the synthesis of viral components. *BMC Genomics* **15**(Suppl 1), S1 (2014).
56. Pettersen, E. F. *et al*. UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
57. Tuttle, L. M. *et al*. Gcn4-Mediator Specificity Is Mediated by a Large and Dynamic Fuzzy Protein-Protein Complex. *Cell Rep.* **22**, 3251–3264 (2018).
58. Tompa, P. & Fuxreiter, M. Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends Biochem. Sci.* **33**, 2–8 (2008).
59. Fung, H. Y. J., Fu, S.-C., Brautigam, C. A. & Chook, Y. M. Structural determinants of nuclear export signal orientation in binding to exportin CRM1. *Elife* **4** (2015).
60. Umunnakwe, C. N., Dorman, K. S., Dobbs, D. & Carpenter, S. Identification of a homogenous structural basis for oligomerization by retroviral Rev-like proteins. *Retrovirology* **14**, 40 (2017).
61. Jackson, P. E., Tebit, D. M., Rekosh, D. & Hammarskjold, M.-L. Rev-RRE Functional Activity Differs Substantially Among Primary HIV-1 Isolates. *AIDS Res. Hum. Retroviruses* **32**, 923–34 (2016).
62. Zapp, M. L., Stern, S. & Green, M. R. Small molecules that selectively block RNA binding of HIV-1 Rev protein inhibit Rev function and viral production. *Cell* **74**, 969–978 (1993).
63. Nakamura, R. L. *et al*. Identification and Optimization of Thienopyridine Carboxamides as Inhibitors of HIV Regulatory Complexes. *Antimicrob. Agents Chemother.* **61**, AAC.02366–16 (2017).
64. Sloan, E. A. *et al*. Limited nucleotide changes in the Rev response element (RRE) during HIV-1 infection alter overall Rev-RRE activity and Rev multimerization. *J. Virol.* **87**, 11173–86 (2013).
65. Powell, D. M., Amaral, M. C., Wu, J. Y., Maniatis, T. & Greene, W. C. HIV Rev-dependent binding of SF2/ASF to the Rev response element: possible role in Rev-mediated inhibition of HIV RNA splicing. *Proc. Natl. Acad. Sci. USA* **94**, 973–8 (1997).
66. D'Agostino, D. M., Felber, B. K., Harrison, J. E. & Pavlakis, G. N. The Rev protein of human immunodeficiency virus type 1 promotes polysomal association and translation of gag/pol and vpu/env mRNAs. *Mol. Cell. Biol.* **12**, 1375–86 (1992).
67. Arrigo, S. J. & Chen, I. S. Rev is necessary for translation but not cytoplasmic accumulation of HIV-1 vif, vpr, and env/vpu 2 RNAs. *Genes Dev.* **5**, 808–19 (1991).
68. Perales, C., Carrasco, L. & González, M. E. Regulation of HIV-1 env mRNA translation by Rev protein. *Biochim. Biophys. Acta - Mol. Cell Res.* **1743**, 169–175 (2005).
69. Grewe, B. *et al*. The HIV-1 Rev Protein Enhances Encapsidation of Unspliced and Spliced, RRE-Containing Lentiviral Vector RNA. *PLoS One* **7**, e48688 (2012).
70. Brandt, S. *et al*. Rev Proteins of Human and Simian Immunodeficiency Virus Enhance RNA Encapsidation. *PLoS Pathog.* **3**, e54 (2007).
71. Blissenbach, M. *et al*. Nuclear RNA export and packaging functions of HIV-1 Rev revisited. *J. Virol.* **84**, 6598–604 (2010).
72. Yedavalli, V. S. R. K., Neuveut, C., Chi, Y.-H., Kleiman, L. & Jeang, K.-T. Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function. *Cell* **119**, 381–92 (2004).
73. Edgcomb, S. P. *et al*. DDX1 Is an RNA-Dependent ATPase Involved in HIV-1 Rev Function and Virus Replication. *J. Mol. Biol.* **415**, 61–74 (2012).
74. Jäger, S. *et al*. Global landscape of HIV-human protein complexes. *Nature* **481**, 365–70 (2012).
75. Yedavalli, V. S. R. K. & Jeang, K.-T. Rev-ing up post-transcriptional HIV-1 RNA expression. *RNA Biol.* **8**, 195–199 (2011).
76. Rueden, C. T. *et al*. ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* **18**, 529 (2017).

Acknowledgements

We thank John Gross, Arabinda Nayak, CJ Umunnakwe, Tyler Faust, and all members of the Frankel lab for helpful comments and review of the manuscript. JDF was supported in part by NIH training grant T32GM007175 and an Amgen Research Excellence Fellowship. SY was supported by the China Scholarship Council. This work was supported by NIH grant P50GM082250 to ADF.

Author Contributions

Conceptualization: B.J., J.D.F. and A.D.F., Investigation: B.J., J.D.F., S.Y. and C.S., Writing: B.J., J.D.F. and A.D.F.; Reviewing & Editing: B.J., J.D.F., S.Y. and A.D.F. B.J. and J.D.F. contributed equally to this manuscript. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-41582-7.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019