ARTICLE

Hsp90 shapes protein and RNA evolution to balance trade-offs between protein stability and aggregation

Ron Geller [1,2], Sebastian Pechmann [1,3], Ashley Acevedo[4,5], Raul Andino[4] & Judith Frydman [1]

Acquisition of mutations is central to evolution; however, the detrimental effects of most mutations on protein folding and stability limit protein evolvability. Molecular chaperones, which suppress aggregation and facilitate polypeptide folding, may alleviate the effects of destabilizing mutations thus promoting sequence diversification. To illuminate how chaperones can influence protein evolution, we examined the effect of reduced activity of the chaperone Hsp90 on poliovirus evolution. We find that Hsp90 offsets evolutionary trade-offs between protein stability and aggregation. Lower chaperone levels favor variants of reduced hydrophobicity and protein aggregation propensity but at a cost to protein stability. Notably, reducing Hsp90 activity also promotes clusters of codon-deoptimized synonymous mutations at inter-domain boundaries, likely to facilitate cotranslational domain folding. Our results reveal how a chaperone can shape the sequence landscape at both the protein and RNA levels to harmonize competing constraints posed by protein stability, aggregation propensity, and translation rate on successful protein biogenesis.

[1] Department of Biology, Stanford University, Stanford, CA 94305, USA. [2] Institute for Integrative Systems Biology (I2SysBio), Universitat de Valencia-CSIC, 46980 Valencia, Spain. [3] Department of Biochemistry, Université de Montréal, Montréal, QC H3T 1J4, Canada. [4] Department of Microbiology and Immunology, UCSF, San Francisco, CA 94131, USA. [5] Present address: Laboratory of Virology and Infectious Diseases, Rockefeller University, Rockefeller University, New York, NY 10065, USA. These authors contributed equally: Ron Geller, Sebastian Pechmann. Correspondence and requests for materials should be addressed to R.A. (email: raul.andino@ucsf.edu) or to J.F. (email: jfrydman@stanford.edu)

A cquisition of mutations is central to evolution but the detrimental effects of most mutations on protein folding and stability limit protein evolvability[1–4]. Molecular chaperones, which suppress aggregation and facilitate polypeptide folding[5], are proposed to promote sequence diversification by buffering against the deleterious effects of destabilizing mutations[6–9]. However, whether and how chaperones control protein evolution remains poorly understood. RNA viruses offer an attractive model to examine such molecular mechanisms of evolution due to their relative simplicity and extreme evolutionary capacity. RNA viral polymerases generate mutations at almost the highest theoretically allowed rates[10,11]. The resulting genetic diversity of viral populations affords rapid adaptation to changing environments[12]. Recent findings link the high population diversity of RNA viruses to pathogenesis, likely due to the need to evolve rapidly in the infected organism[13]. As viral proteins meet folding and stability challenges akin to those of host proteins, and utilize the host-cell translation and folding machineries, viruses offer a unique opportunity to examine how chaperones shape protein evolution.

Poliovirus is a non-enveloped enterovirus that replicates in the cytosol of human cells. Infection leads to shutoff of host protein synthesis; thus the host translation and protein folding machineries are entirely dedicated to produce viral proteins[14]. Hsp90, an abundant ATP-dependent chaperone that facilitates the folding and maturation of many cellular metastable proteins[15–17], only interacts with a single poliovirus protein, the P1 capsid precursor[18]. Hsp90 activity is required for P1 folding and subsequent proteolytic maturation into capsid proteins VP0, VP1, and VP3[18]. Importantly, P1 folding is the only process in the poliovirus replication cycle that requires Hsp90. Pharmacological inhibition of Hsp90 activity with the highly selective Hsp90 inhibitor Geldanamycin[19] (GA) specifically impairs P1 folding and maturation but does not affect viral protein translation nor the folding or function of other viral proteins[18]. The high diversity and evolutionary capacity of poliovirus, together with the fact that it harbors only a single protein that requires Hsp90 for folding, the capsid precursor P1, provide an ideal system to examine the role of Hsp90 in protein evolution (Fig. 1a).

Herein, we capitalize on this system to address how Hsp90 influences the evolution of the poliovirus P1 protein. We find that Hsp90 can modulate the fitness of variants in P1 under normal conditions, selecting against particular mutations. Moreover, we observe that under conditions of reduced chaperone function, poliovirus populations harbor more variants that destabilize P1 and less variants that increase aggregation propensity and hydrophobicity. Hence, Hsp90 appears to mediate a trade-off between key conflicting aspects of protein folding, stability, and aggregation propensity. Strikingly, we observe clusters of deoptimized synonymous codons at domain boundaries under conditions of low Hsp90 activity, linking chaperone function with translation kinetics, and cotranslational folding. Our results reveal that a key eukaryotic chaperone influences the sequence landscape of its client protein at both the protein and RNA levels, mediating the competing constraints posed by protein stability, aggregation propensity, and translation rate.

Results

Hsp90 influences the repertoire of antibody escape mutants.
During infection, viral capsids are subject to intense selective pressure by the host immune system[20]. Viral protein diversity is the key to virus survival as it creates a reservoir of capsid variants enabling escape from neutralization by circulating antibodies. To assess whether Hsp90 influences viral capsid evolution, we examined the spectrum of neutralizing antibody escape variants

in viral populations that were generated under normal conditions ($Hsp90_N$) or following Hsp90 inhibition by treatment of infected cells with Geldanamycin (GA) ($hsp90_i$; Fig. 1b). We speculated that if Hsp90 has a role in influencing the evolution of its client proteins, we would observe a difference in the type or frequency of mutations giving rise to antibody resistance between $Hsp90_N$ and $hsp90_i$ populations. Two neutralizing monoclonal antibodies (mAb423 and mAb427) targeting different epitopes in the poliovirus capsid[21] were employed, followed by sequencing across their respective epitope. Escape variants conferring resistance to neutralization were identified in both conditions, in two independent experiments (Fig. 1b). For mAb423, resistance in the $Hsp90_N$ population mapped in both replicates to several different substitutions in two P1 sites: K402 and R412 (Fig. 1c, Supplementary Fig. 1a, Supplementary Table 1). In contrast, mAb423 resistance in the $hsp90_i$ population was dominated by a single substitution in both replicates, K402E ($p < 0.0001$ by Fisher's test; Fig. 1c, Supplementary Fig. 1a, Supplementary Table 1). The R412W mutation, often encountered in the $Hsp90_N$ population, was significantly disfavored in the $hsp90_i$ population ($p < 0.0001$ by Fisher's test; Fig. 1c, Supplementary Fig. 1a, Supplementary Table 1).

For the other antibody, mAb427, escape neutralization variants observed in both $Hsp90_N$ and $hsp90_i$ populations were distributed across multiple sites in the epitope. Several distinct variants at position 235 (N235) were significantly enriched in the $Hsp90_N$ population in both replicates ($p = 0.0014$ by Fisher's test; Fig. 1d, Supplementary Fig. 1b, Supplementary Table 2). In contrast, mutations at positions P239 and N216 were observed at high frequency in the $hsp90_i$ population in replicate 1 and 2, respectively, but were nearly absent from the $Hsp90_N$ population ($p < 0.05$ by Fisher's test for both N216 and P239; Fig. 1d, Supplementary Fig. 1b, Supplementary Table 2). These results suggest that Hsp90 influences the types of P1 mutations that can exist in poliovirus populations (P1 sequence space) by selecting both for and against specific sequence variants.

Hsp90 modulates the fitness of mutations in its client protein.
The near absence from the $Hsp90_N$ population of mAb427 escape variants at sites N216 and P239 suggests that normal Hsp90 levels reduce the fitness of these sequence variants. To directly test the notion that Hsp90 negatively selects certain sequence variants, we assessed the relative fitness of viruses harboring either the $hsp90_i$ selected variants at P1 sites 216 and 239, or the $Hsp90_N$ variants observed under Hsp90-normal conditions. Viruses encoding the $Hsp90_N$ sequences N216 and P239 were competed with viruses encoding the $hsp90_i$ variants K216 or S239, under conditions of either normal or low Hsp90 activity (Fig. 1e). A mix with an initial ratio of 1:5 $Hsp90_N$ to $hsp90_i$ variant was serially passaged ten times with or without Hsp90 inhibitor in two independent experiments. Sequence analysis of passage 10 populations (Fig. 1e, Supplementary Fig. 2) showed that under normal Hsp90 conditions, the $Hsp90_N$ variants N216 and P239 were strongly selected for in both independent experiments, completely dominating the population after 10 passages despite the initial excess of $hsp90_i$ virus (Fig. 1e, normal Hsp90). In contrast, when passaged in the presence of Hsp90 inhibitor (Fig. 1e, low Hsp90), $hsp90_i$ variant K216 was present in ~50% of the population (Fig. 1e, Supplementary Fig. 2a) and variant S239 completely dominated the population (Fig. 1e, Supplementary Fig. 2b). Importantly, the $hsp90_i$ variants K216 and S239 do not show resistance to Hsp90 inhibition (Supplementary Fig. 3), in agreement with previous findings that Hsp90 inhibitors are refractory to the development of drug resistance[18]. These findings demonstrate that Hsp90 levels influences the fitness of particular sequence variants in a
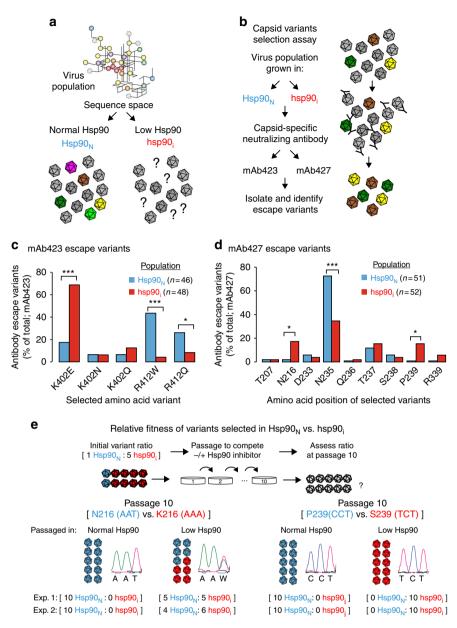
Fig. 1 Hsp90 activity influences viral capsid fitness. **a** Does Hsp90 activity influence the sequence space of its client, poliovirus capsid protein P1? The schematic indicates a theoretical sequence space for P1 in viral populations, with each branch indicating two sequence variants linked by a point mutation. $Hsp90_N$ and $hsp90_i$ represent viral populations generated under normal Hsp90 activity, or under reduced Hsp90 activity by treatment with Geldanamycin (GA). WT capsids are indicated in gray, mutant capsids in color, and question marks indicate unknown capsid mutants. **b** Outline of capsid variants selection assay using neutralizing antibodies mAb423 and mAb427. **c**, **d** Abundance and distribution of mAb423 (**c**) or mAb427 (**d**) escape variants identified in $Hsp90_N$ and $hsp90_i$ populations. Each graph represents the sum of all mAb-resistant variants identified in two independent experiments (see Supplementary Fig. 1), with the total number of mutants (n) indicated. **e** Evaluation of the relative fitness of the $hsp90_i$-enriched variants N216K and P239S identified in **d** under normal or Hsp90-inhibited conditions. An initial infection with a 1:5 ratio of $Hsp90_N$ (N216 or P239) to $hsp90_i$ variant (K216 or S239) virus was used, and viruses serially passaged under normal or Hsp90-inhibited conditions for 10 passages. The results of the competition assays were assessed by Sanger sequencing of the population, and is shown for passage 10. Significance: *$p < 0.05$, **$p < 0.01$, ***$p < 0.005$ by two-tail Fisher's exact test after multiple test correction using the false discovery rate (FDR) method

client protein. Given the growing interest in Hsp90 inhibitors as broad-spectrum antivirals[18,22,23], the effect of Hsp90 on the fitness of escape mutants from immune selection may have important therapeutic implications.

Hsp90 influences the diversity of poliovirus populations. To gain a better understanding of how Hsp90 shapes virus evolution, we next examined the diversity of poliovirus populations at unprecedented resolution using recent technological

advancements in ultra-deep sequencing that provide a detailed description of the viral population mutation composition[24]. We considered three possible scenarios by which Hsp90 influences the functional sequence space of a protein (Fig. 2a). The first possibility, that Hsp90 has no role shaping sequence space (no-effect; Fig. 2a, left panel), is negated by the experiments in Fig. 1. The second model stems from the role of Hsp90 and other chaperones in buffering destabilizing mutations[9,25–30] (buffers mutations; Fig. 2a, middle panel). This buffering model, which remains controversial[31], predicts that Hsp90 inhibition will
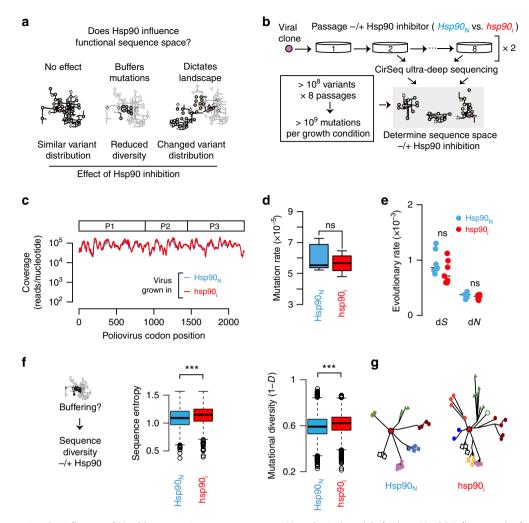
Fig. 2 Globally assessing the influence of Hsp90 on protein sequence space. **a** Hypothetical models for how Hsp90 influences the functional sequence space of poliovirus: no-effect (left), buffers mutations (center), or dictates landscape (right). Schemes illustrate the respective predicted outcome of Hsp90 inhibition on sequence space for each model. Gray lines indicate available mutational networks while black lines indicated those that are present under the different proposed scenarios. **b** Experimental strategy to examine the effects of Hsp90 inhibition on poliovirus sequence space. Poliovirus was propagated over eight passages under normal or low Hsp90 conditions via GA treatment. Population samples containing >10^8 variants each were collected from passages 2–8 and subjected to CirSeq deep sequencing and analysis. **c** Average sequence coverage obtained with CirSeq across the poliovirus genome for all populations $Hsp90_N$ (blue trace) and $hsp90_i$ (red traces). **d** Mutation rates for the full virus genome in $Hsp90_N$ and $hsp90_i$ populations. **e** Evolutionary rates for P1 in $Hsp90_N$ and $hsp90_i$ populations (full genome in Supplementary Fig. 4c). **f** Does Hsp90 buffering lead to reduced P1 sequence diversity in $hsp90_i$ populations? Sequence entropy (left panel) and mutational diversity (right panel) for P1 coding region is higher for $hsp90_i$ than for $Hsp90_N$ populations analyzed by CirSeq. **g** Representation of distinct sequence diversity for P1 from $Hsp90_N$ and $hsp90_i$ virus populations. The original master sequence (red, middle circle) gives rise to variants (different colors and symbols) through mutation. For boxplots, the center line represents the median, the bound box the interquartile range, and the whiskers 1.5× the interquartile range. Significance: ns, not significant ($p > 0.05$); ***$p < 0.005$ by the Mann–Whitney–Wilcoxon test

subject less stable mutations to purifying selection due to the reduced fitness of these mutants in the absence of the chaperone[32,33]. This should favor a smaller sub-population of stable variants, resulting in reduced population diversity. Finally, a third possible model is that Hsp90 shapes the sequence space of its client proteins in a manner that has yet to be appreciated (dictates landscape; Fig. 2a, right panel).

Starting with a single viral clone, we sampled virus populations from eight different passages under conditions of normal Hsp90 activity ($Hsp90_N$) or low Hsp90 activity ($hsp90_i$) by treatment with the Hsp90 inhibitor GA[19] in two independent replicates (Fig. 2b). We then employed high fidelity CirSeq ultra-deep sequencing[24] to define the sequence composition of these poliovirus populations (Fig. 2b). Extensive sequence coverage was obtained for both replicas, with most sites in the coding sequence covered over 100,000 times (Fig. 2c, Supplementary

Fig. 4a). The consensus sequence was unchanged in all samples, as previously observed upon passaging poliovirus in the presence of Hsp90 inhibitors[18]. Furthermore, no global differences in the overall mutation rates of the $Hsp90_N$ and $hsp90_i$ virus populations were observed (Fig. 2d, $p = 0.71$ by Mann–Whitney–Wilcoxon (MWW) test, and Supplementary Fig. 4b). Similarly, no significant differences in overall evolutionary rates were observed for both synonymous (dS) and non-synonymous mutations (dN) as a function of Hsp90 activity level, either for Hsp90-client protein P1, or across the entire genome (Fig. 2e, dS: $p = 0.21$ and dN: $p = 0.25$ by MWW test; and Supplementary Fig. 4c). Thus, global evolutionary rates were not affected by Hsp90 activity, consistent with previous findings that Hsp90 inhibition does not affect the viral polymerase[18]. Interestingly however, protein mutational diversity in P1 was significantly higher in the $hsp90_i$ condition, in contrast to the expectation from the buffering model

(Fig. 2f). Two alternative metrics, sequence entropy (Fig. 2f, $p < 8 \times 10^{-26}$ by MWW test; and Supplementary Fig. 4d) and mutational diversity $(1 - D)$ with D as Simpson's index of diversity (Fig. 2f, $p < 8.7 \times 10^{-21}$ by MWW test; and Supplementary Fig. 4d) both indicated that reducing Hsp90 activity enhances population diversity (Fig. 2g). Together with the data from Fig. 1, these analyses suggest that Hsp90 influences the sequence landscape within the population (Fig. 2a, right panel).

Hsp90 influences the capsid sequence space. Molecular chaperones assist protein homeostasis by promoting folding upon translation, maintaining stability, and also preventing protein aggregation[5,34,35] (Fig. 3a). These processes are particularly relevant to viral capsids, which must be produced in high levels and assembled from numerous subunits, while avoiding aggregation[36,37]. Furthermore, once assembled, capsids must be sufficiently stable to protect the viral genome in harsh environments. To further understand the role of Hsp90 in shaping protein evolution, we compared the viral populations generated under normal ($Hsp90_N$) or Hsp90-inhibited ($hsp90_i$) conditions to identify variants that differed significantly in their effect on protein stability or aggregation propensity across passages (Fig. 3b). First, we used the experimentally-rooted and well-validated FoldX approach[38–40] to calculate the effects of all possible single point mutations on protein stability in the three viral proteins for which crystal structures are available, the capsid P1, the protease 3C and the RNA polymerase 3D (FoldX algorithm[38] see Methods and Supplementary Fig. 5a). We then compared $Hsp90_N$ and $hsp90_i$ populations to identify P1 sites that significantly differ in their destabilizing variant frequencies across passages (Site$_{destab}$). Next, we calculated the effect of mutations on sequence aggregation propensity employing the widely used TANGO algorithm[41–43] (see Methods and Supplementary Fig. 5b) in these same viral proteins to uncover P1 sites that differ in their ability to tolerate variants of increased aggregation

propensity (Sites$_{agg}$) between the $Hsp90_N$ and $hsp90_i$ population. For each of these analyses, we employed a robust statistical approach that examines the differences in the ability of each codon position to consistently accommodate destabilizing or aggregation-prone mutations across all passages (see Methods). Surprisingly, there were significantly more Site$_{destab}$ in the Hsp90-client P1 under low Hsp90 activity (Fig. 3c; $p = 0.0009$ by Cochran–Mantel–Haenszel (CMH) test; Supplementary Fig. 5d). Conversely, the number of Sites$_{agg}$ in P1 was significantly reduced under low Hsp90 activity (Fig. 3d, $p = 0.00036$ by CMH test; Supplementary Fig. 5e). The incidence of Sites$_{agg}$ and Site$_{destab}$ in the non-Hsp90 dependent viral proteins 3C and 3D was relatively unaffected by Hsp90 inhibition (Fig. 3c, d: $p = 0.21$, 3D: $p = 1$ by CMH test). These analyses indicate that Hsp90 preferentially affects the sequence space of its client P1. Paradoxically, unlike what we expected from the 'buffering' model, reducing Hsp90 activity increases the incidence of variants harboring destabilizing mutations as well as variants that reduce aggregation propensity, in support of a 'dictating landscape' model.

We next examined the distribution of Sites$_{agg}$ and Site$_{destab}$ with respect to the folded P1 structure, namely whether they map to buried (Brd), surface-exposed (Sfc), or interface (Int) positions (Fig. 3e). For both Sites$_{agg}$ and Site$_{destab}$, major differences between $Hsp90_N$ and $hsp90_i$ mapped to buried (Brd) positions within the structure of the viral capsid (Fig. 3e and Supplementary Fig. 5d, e). Under low Hsp90 conditions, buried (Brd) destabilizing variants were increased (Site$_{destab}$; $p = 0.006$ by Fisher's test) and aggregation-prone sites decreased (Sites$_{agg}$; $p < 0.004$ by Fisher's test) compared to surface-exposed (Fig. 3e; Sfc) or interface regions (Fig. 3e; Int). As buried regions are primarily exposed in non-native conformations, such as those generated during cotranslational folding, these findings indicate that Hsp90 activity may be critical to prevent aggregation during initial protein folding, rather than during assembly, for which interface residues are likely of greater relevance.
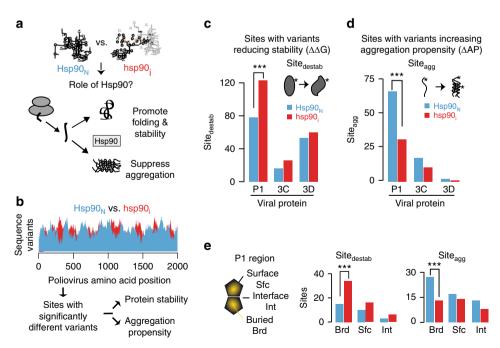

Fig. 3 Hsp90 influences the capsid sequence space. **a** Role of Hsp90 in pathways of protein folding and aggregation. **b** Analysis pipeline to identify codon sites that differ significantly in their ability to support destabilizing variants (Site$_{destab}$) or aggregation-prone (Site$_{agg}$) variants between $Hsp90_N$ and $hsp90_i$ populations. Significant sites were identified as described in Methods. **c, d** Number of Site$_{destab}$ (**c**) and Site$_{agg}$ (**d**) observed in viral proteins: P1 capsid protein; 3C protease; and 3D polymerase. **e** Distribution of Site$_{destab}$ and Site$_{agg}$ at buried (Brd), interface (Int), or surface (Sfc) positions across folded P1 (based on crystal structure 2PLV). Significance: ***$p < 0.005$ by two-tail Fisher's exact test

Hsp90 supports variants of increased hydrophobicity.
Disfavoring aggregation-prone variants should be beneficial for the virus, while favoring destabilizing mutations should be detrimental. How can the contradictory effects of low chaperone activity on these two key protein-folding properties be reconciled? We reasoned that hydrophobicity, a driving force in protein folding that influences both protein stability and aggregation[44,45] may explain these effects (Fig. 4a). Enhancing buried hydrophobicity can increase the stability of the folded core, but can also increase aggregation propensity. We thus examined if $Hsp90_N$ and $hsp90_i$ populations differ in their ability to accommodate sites with enhanced hydrophobicity (Sites$_{hydro}$; Fig. 4b). Strikingly, inhibiting Hsp90 significantly reduced the number of Sites$_{hydro}$ in P1 and to a lesser extent also in non-Hsp90 clients 3C and 3D (Fig. 4b). The effect on P1 was dramatic for both replicas (Fig. 4b, $p = 5 \times 10^{-9}$ by CMH test, and Supplementary Fig. 5f) and observed both in buried, surface-exposed and interface sites

(Supplementary Fig. 5c). These results indicate that Hsp90 enables sequence variants of increased hydrophobicity to survive in the population, which can promote protein stability but also increase aggregation propensity.

To better understand the surprising finding that reduced chaperone function results in an increase of destabilizing variants, we next examined the properties of individual Site$_{destab}$ variants in $Hsp90_N$ and $hsp90_i$ viral populations. Site$_{destab}$ variants in P1 were examined with respect to their ability to change aggregation propensity (ΔAP) and hydrophobicity (ΔHyd) (Fig. 4c, d). Destabilizing variants observed under normal Hsp90 conditions ($Hsp90_N$) were significantly more aggregation-prone than destabilizing variants present at reduced chaperone activity ($hsp90_i$; Fig. 4c, d; $p = 0.002$ by MWW test, and Supplementary Fig. 5g). Furthermore, Site$_{destab}$ variants in $hsp90_i$ were significantly less hydrophobic than in $Hsp90_N$ or those expected by chance (i.e., randomly selected mutations; Fig. 4c, $p < 10^{-7}$ by
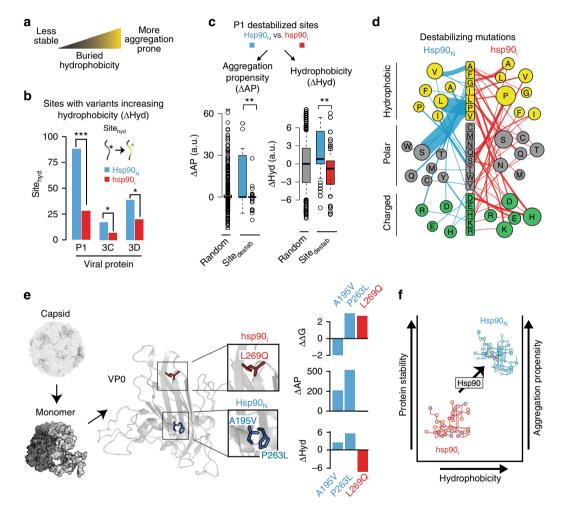

Fig. 4 Hsp90 mediates a trade-off between stability and aggregation propensity. **a** Effect of buried hydrophobicity on protein stability and aggregation propensity. **b** Sites with increased hydrophobic variants (Site$_{hydro}$) in $Hsp90_N$ and $hsp90_i$ populations. **c** Properties of P1 destabilized sites (Site$_{destab}$) from $Hsp90_N$ and $hsp90_i$: aggregation propensity (left panel) and hydrophobicity (right panel). **d** Distinct nature of destabilizing variants (Site$_{destab}$) in P1 of $Hsp90_N$ and $hsp90_i$ populations. Original amino acid is indicated in the middle. Color indicates side-chain properties: Yellow, hydrophobic; gray, polar; green, charged. Mutations to specific residues in $Hsp90_N$ (left) and $hsp90_i$ (right) are indicated, where circle size indicates mutation occurrence. The width of lines connecting each mutation represents the number of occurrences for each variant. **e** Exemplar P1 variants enriched in $Hsp90_N$ and $hsp90_i$ populations. Location in P1 folded structure and effects on stability, aggregation propensity, and hydrophobicity are indicated. Mutations A195V and P263L (blue) observed in $Hsp90_N$, and L269Q (red) in $hsp90_i$. Structure from 2PLV, with VP2 monomer, and assembled capsid indicated. **f** Schematic effect of Hsp90 on sequence space. Hsp90 allows variants to explore a region in sequence space of increased hydrophobicity and aggregation propensity, but also of increased stability. Lower Hsp90 constrains sequence space variants to a region of reduced stability, hydrophobicity, and aggregation propensity. For boxplots, the center line represents the median, the bound box the interquartile range, and the whiskers 1.5× the interquartile range. Significance: $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.005$ by two-tail Fisher's exact test or Mann–Whitney–Wilcoxon test

MWW test, and Supplementary Fig. 5g). This indicates that $Hsp90_N$ and $hsp90_i$ accommodate different kinds of destabilizing mutations. Thus, destabilizing mutations in $Hsp90_N$ have significantly increased sequence hydrophobicity and aggregation propensity compared to that expected by chance (Fig. 4c, left panel, $p < 10^{-5}$ by MWW test and Supplementary Fig. 5g). In contrast, destabilizing mutations in $hsp90_i$ have significantly decreased sequence hydrophobicity and aggregation propensity (Fig. 4c, right panel, $p = 0.0007$ by MWW test, and Supplementary Fig. 5g). More specifically, many destabilizing mutations in the $Hsp90_N$ population resulted from the loss of proline residues (Fig. 4d, Supplementary Fig. 5h). In contrast, destabilizing variants in the $hsp90_i$ population were more diverse and reduced hydrophobicity by mutation to polar or charged residues (Fig. 4d, Supplementary Fig. 5h). We conclude that Hsp90 enables accumulation of destabilizing variants with increased aggregation propensity and hydrophobicity, while reducing Hsp90 activity promotes destabilizing variants that reduce hydrophobicity and aggregation propensity.

The interplay between Hsp90 action and individual mutations affecting protein stability, aggregation propensity and hydrophobicity is illustrated by a set of variants found at the interface between core beta strands in VP0, the N-terminal folding unit of P1 (Fig. 4e). For instance, a stabilizing mutation found in $Hsp90_N$, A195V, should be beneficial to protein folding and stability; however, this mutation also increases P1 aggregation propensity and hydrophobicity, and is thus strongly selected against in the $hsp90_i$ population ($p < 1 \times 10^{-4}$ by CMH test). Another $Hsp90_N$ variant, P263L, also increases aggregation propensity and hydrophobicity but, in this case, reduces protein stability. This very detrimental mutation is strongly depleted under $hsp90_i$ conditions ($p < 1 \times 10^{-4}$ by CMH test). In contrast, L269Q, which reduces sequence hydrophobicity and does not alter aggregation propensity, is favored in $hsp90_i$, despite being a strongly destabilizing mutation ($p < 10^{-10}$ by CMH test). These analyses further illustrate that, for poliovirus P1 capsid protein, the dominant selective force under low Hsp90 activity is the reduction of hydrophobicity and aggregation propensity, even at a cost to protein stability.
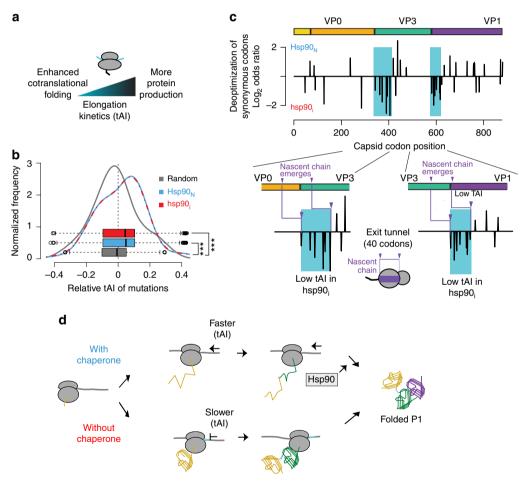

Fig. 5 Localized codon de-optimization upon adaptation to Hsp90 inhibition. **a** Interplay between the rate of translation, cotranslational folding, and protein production. **b** Both $Hsp90_N$ and $hsp90_i$ poliovirus populations accumulate synonymous variants resulting in global codon optimization. The change in codon optimality was calculated for all synonymous mutations observed in $hsp90_i$ and $Hsp90_N$ populations as well as for a random sample of mutations (random) based on the HeLa cell tRNA adaptation index (tAI). **c** Distribution of sites harboring significantly deoptimized synonymous codons across P1 in $hsp90_i$ and $Hsp90_N$ populations. The domain structure of P1 is indicated. Insets show the two clusters of deoptimized codons in $hsp90_i$ (indicated with blue shading) relative to domain boundaries of VP0/VP3 (left) and VP3/VP1 (right). The approximate size of the ribosomal exit tunnel and the point at which the nascent-chain (n. chain) emerged from the ribosome are indicated in purple. **d** Model of the potential impact of chaperone levels on codon de-optimization at domain boundaries. Under conditions of Hsp90 inhibition, the virus locally reduces translation speeds to promote cotranslational folding. While enhancing correct cotranslational P1 folding, this strategy may slow protein production. For boxplots, the center line represents the median, the bound box the interquartile range, and the whiskers 1.5× the interquartile range. Significance: ***$p < 0.005$ by Mann–Whitney–Wilcoxon test

Our analysis exposes the role of Hsp90 in balancing an intrinsic conflict between stability and aggregation, shaping protein evolution (Fig. 4f). While increased hydrophobicity benefits the stability of the folded state, it also increases the aggregation propensity of folding intermediates, such as those generated during protein synthesis. Thus, chaperone dependence, which reduces aggregation in vivo, may allow exploration of regions within sequence space leading to more stable, and thus more versatile, proteins (Fig. 4f). This is consistent with the high aggregation propensity of obligate chaperone substrates in the cell[46,47]. Importantly, by favoring more stabilizing but also more aggregation-prone sequence variants, chaperones appear to restrict the sequence space and render proteins vulnerable to aggregation under conditions of chaperone impairment, such as cellular stress or aging.

Hsp90 activity can influence synonymous codon choice. Our analysis points to the importance of Hsp90 in preventing aggregation during initial protein folding (Fig. 3e). The vectorial nature of translation renders proteins particularly susceptible to aggregation[48], since emerging domains that cannot yet fold will expose hydrophobic elements to the crowded cellular milieu[49,50]. Importantly, chaperones have a key role protecting nascent chains and facilitating their folding[51]. Since the speed of translation is proposed to be directly attuned to optimize cotranslational folding and processing of nascent chains[52–55], we hypothesized that virus adaptation to low Hsp90 levels may involve optimization of the translation elongation rate. Some tRNAs are much more abundant than others and codon choice can have important effects on translation kinetics[56–58] (Fig. 5a). Accordingly, viruses have been found under direct selection for their synonymous codon usage[59] whose deregulation often brings direct fitness consequences[60–62]. We thus examined whether reducing Hsp90 levels affected the synonymous choice of codons in P1 using the established tRNA adaptation index[63] (tAI) for both the human genome and for HeLa cells used for virus passaging. Of note, the tAI has been shown to correlate with in vivo translation speed[57]. Overall, both $Hsp90_N$ and $hsp90_i$ populations demonstrate a clear trend to generally optimize overall codon usage to match that of the host-cell (Fig. 5b; $Hsp90_N$: $p < 10^{-6}$ by MWW test; $hsp90_i$: $p < 10^{-6}$; Supplementary Fig. 6a,b). This indicates that poliovirus adapts to the host-cell to enhance overall protein production, establishing a direct link between viral infection and adaptation to the host-cell tRNA pool.

However, under conditions of Hsp90 inhibition, we also observed clusters of sites harboring synonymous mutations that deoptimized codons at specific locations along the P1 coding sequence (Fig. 5c and Supplementary Fig. 6c, d). Specifically, there were two major clusters of deoptimized codons in the $hsp90_i$ population (Fig. 5c, highlighted in blue). Each cluster mapped to the N-terminal coding region of VP3 and VP1, respectively. The P1 polyprotein encompasses three domains that, upon completion of folding, are processed proteolytically to generate processed capsid subunits VP0, VP3, and VP1[37]. Interestingly, accounting for the length of the ribosomal exit tunnel, the deoptimized codon clusters were positioned to promote a local slow-down of translation elongation upon emergence of the complete VP0 and VP3 folding domains from the ribosomal exit site (Fig. 5c, bottom panels). Importantly, pulse chase experiments show that Hsp90 inhibition does not affect global translation speeds in poliovirus infected cells[18], supporting the idea that synonymous changes affect the translation rate locally. We propose that synonymous mutations are selected at lower Hsp90 activity to locally reduce translation speed at inter-domain boundaries to favor cotranslational folding. Consistent with this notion, the only

member of the picornavirus family whose P1 protein does not require Hsp90 for folding, Hepatitis A virus, relies on highly deoptimized codon usage for the capsid coding region for fitness[64]. These findings suggest that achieving overall higher translation rates also requires a trade-off with efficient cotranslational folding, which can be balanced by chaperone action or in its absence, by selective codon de-optimization at specific sites (Fig. 5d).

Discussion

Although clearly essential for protein folding[15], the role of Hsp90 in evolution has remained unclear[31]. Our study demonstrates that the chaperone Hsp90 helps shape the evolutionary path of a viral protein at the polypeptide and RNA levels. Hsp90 alters the protein sequence space by modulating trade-offs between protein stability, aggregation and hydrophobicity at the polypeptide level, and the trade-offs between translation speed and cotranslational folding at the RNA level. Hsp90 allows exploration of variants that increase aggregation propensity and hydrophobicity, which may also lead to increased protein stability (Fig. 6). In low Hsp90, the viral sequence space expands to explore variants that reduce aggregation propensity at the expense of stability.

Our experiments provide insights into how chaperones may have shaped protein evolution by modulating the energy landscape of protein folding. Upon synthesis, many globular proteins populate non-native states as well as aggregation-prone-intermediates; of note, aggregates are often more stable than the native folded state[65] (Fig. 6). Chaperones guide the folding polypeptide through this landscape to prevent aggregation and favor the energetically stable folded conformation. Without chaperones, alternative solutions within sequence space must be explored that alleviate the pressure to reduce intrinsic aggregation propensity; this may preclude the emergence of more hydrophobic variants that stabilize the native state. These considerations may be particularly relevant in the crowded environment of the cell, as well as for topologically complex proteins such as viral capsids, which must assemble elaborate oligomeric structures[66]. However, the intrinsic conflict between protein stability and
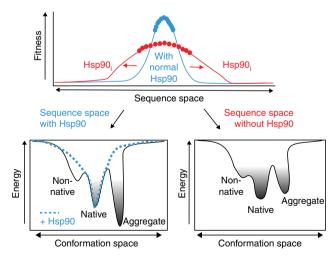

Fig. 6 Hsp90 modulates the energy landscape of its client protein. Role of Hsp90 in shaping protein evolution by modulating the energy landscape of protein folding and balancing trade-offs between protein stability, aggregation propensity, and folding. Hsp90 allows exploration of variants that increase aggregation propensity and hydrophobicity, which may also lead to increased protein stability. Under reduced Hsp90 conditions, the viral sequence space expands to explore variants that reduce aggregation propensity at the expense of stability. See text for details

aggregation propensity is a property of all proteins[65] and we speculate that our results have general implications for further understanding how chaperones shaped the evolution of proteins.

Our experiments also suggest that chaperone activity influences the RNA sequence space. Low Hsp90 activity promotes deoptimizing codons following translation of domains. Such modulation of local translation kinetics may enhance cotranslational folding by allowing more time for early folding intermediates to adopt correct conformations, avoid misfolding and aggregation, or by allowing more time for recruiting and interacting with chaperones. These findings highlight the close relationship between protein and RNA sequence space, and show that the evolutionary trajectory of both may be shaped by chaperones.

Not all proteins require chaperone assistance for folding[67]. Our results suggest that one benefit of acquiring chaperone dependence is to resolve evolutionary trade-offs that allow proteins to be translated faster and to occupy specific regions of sequence space. Chaperone dependence thus likely supported acquisition of complex folds that expand protein function, including metastable structures[17,68], multimeric assemblies[5,69], and multi-domain architectures[5,69]. On the other hand, chaperone substrates are constrained by chaperone-specific preferences and become sensitive to the levels and activity of that chaperone in different cells and environmental conditions. Chaperone activity can vary with stress, cell type, and even age[34], and these changes may yield variance in the folding and activity of a chaperone-dependent protein. In the particular case of viral proteins, where survival and adaptation is directly dependent on population diversity[13], chaperones likely have a fundamental role in virus protein evolution, adaptation to specific environments, like tissues and stress conditions, and ultimately pathogenesis.

Methods

Strains and reagents. All experiments were conducted in HeLa S3 cells (ATCC, CCL-2.2) cultured in DMEM/F12 (Invitrogen) containing 10% FCS, penicillin and streptomycin and at 37 °C. The Mahoney type 1 (WT) and Sabin type 1 strains of poliovirus were generated by electroporation of in vitro transcribed RNA into Hela S3 cells[18]. For viral infections, cells were incubated with the virus for 30 min at 37 °C, after which media was added and infection allowed to continue until all cells were dead. Subsequently, two freeze-thaw cycles were performed to release intra-cellular viruses, cellular debris were removed by low speed centrifugation, and virus production was assessed by plaque assay. Geldanamycin (GA; LC labs) was diluted in DMSO and used at the indicated concentration. Monoclonal poliovirus antibodies were generously provided by Dr. Morag Ferguson at NIBSC, Blanche Lane, South Mimms, Potters Bar, Hertfordshire.

Antibody neutralizations and competition assays. For selection of antibody escape mutants, ~10^5 (mAb427) or 10^6–10^7 (mAb423) plaque forming units (PFU) from virus populations generated in the absence or presence of 1 µM GA were mixed with 1 µl of antibody for 1 h at 25 °C and then used to infect confluent cells in 6 cm plates for 30 min. Cells were then washed and overlaid with DMEM/F12 containing 1% agar and 1:2000 dilution of antibody. Twenty-four to 30 h later, plaques were picked and amplified by inoculating onto fresh cells in 12 well plates. Upon CPE, viral RNA was extracted by Trizol or ZR Viral RNA kit (Zymo Research), RT-PCR performed with Thermoscript reverse transcriptase (Invitrogen) using random primers, and PCR performed using primers F676 (5′-GCTCCATTGAGTGTGTTTACTCTA-3′) and R3463 (5′-ATCATCCTGAGT GGCCAAGT-3′) and PfuUltra II Fusion HS (Stratagene). For competition experiments, 10^5 PFU of mutants were mixed with 2×10^4 pfu of WT poliovirus and passaged as above. p-values were calculated using a two-tailed Fisher's test and multiple testing correction using the false discovery rate (FDR) method[70].

CirSeq sequencing. CirSeq was performed according to published protocols[24,71]. Briefly, Hela S3 cells were serially passaged at a multiplicity of infection (m.o.i.) of 0.1 with 10^6 PFU for 8 h (one replication cycle) in the presence ($hsp90_i$) or absence ($Hsp90_N$) of 0.3 μM Hsp90 inhibitor Geldanamycin. Each passage was amplified in Hela S3 cells at high m.o.i. and total cellular RNA was extracted at 6 h post infection. Viral RNA was enriched by poly(A) purification, fragmented, circularized and reverse transcribed to generate cDNAs containing tandemly repeated viral sequences. Sequencing libraries were generated from cDNA by second strand synthesis, end repair, dA-tailing and ligation of Illumina TruSeq indexed adaptors to the DNA ends. Sequencing was performed on a HiSeq2500

using HiSeq Rapid SBS V2 reagents and 300–310 bp single-end reads with indexing as appropriate. Reads were processed with the CirSeq software[24,71], available at https://github.com/ashleyacevedo/CirSeq, to generate a consensus sequence of tandem repeats within each read. Duplicate reads with identical break points were filtered, and aligned reads converted to codon counts at each position in the coding region.

Cirseq analysis. The virus mutation rate was computed by analyzing the rate of nonsense mutations to stop codons[72]. The evolutionary rates dN and dS were computed with established count model of codon sequence evolution by Nei and Gojobori[73]. The notation of dN and dS was chosen to reflect our focus on the molecular evolution of poliovirus proteins even though we analyze virus populations[74]. To evaluate mutational diversity, two independent measures that quantify diversity were used. First, Shannon's Entropy, a metric from information theory that quantifies information content, was computed of the mutational space per position as

$$H = -\sum_i p_i \log(p_i), \tag{1}$$

where p_i is the probability of observing mutation i among the mutations. Second, Simpson's Index of diversity $1 - D$, a biodiversity measure from ecology, was used to assess mutational diversity per position as

$$D = \sum_i \left(\frac{n_i}{N}\right)^2, \tag{2}$$

where n is the number of mutations of each type i, and N is the total number of all mutations. For both metrics, diversity profiles of the poliovirus coding sequence, considering all sites with at least 1 mutation, were computed for $Hsp90_N$ and $hsp90_i$. Upon validation that the observed differences between $Hsp90_N$ and $hsp90_i$ are consistent for all passages, average profiles across passages are reported. Global differences in mutation rate, evolutionary rates, and mutational diversity were tested for statistical significance with the Wilcoxon–Mann–Whitney test.

To test how host-cell chaperone levels influence mutations in the virus populations that impact protein folding and stability, we considered three classes of mutations namely destabilizing mutations, mutations that increase the protein aggregation propensity, and mutations that significantly alter sequence hydrophobicity. The effect of all amino acid mutations on protein stability was predicted with FoldX[38] (foldxsuite.crg.es) for poliovirus proteins P1 (PDB structure 2PLV), 3C (4DCD), and 3D (4NLR). The effect of amino acid mutations on protein aggregation propensity was predicted with Tango[41] (tango.crg.es), and changes in amino acid hydrophobicity upon mutation were assessed by the Kyte–Doolittle scale normalized to zero-mean and unitary standard deviation[75]. Mutations were subsequently classified into effect (E) and no-effect (NE), i.e., destabilizing if the predicted stability effect ΔΔG > 1 kcal/mol, and otherwise non-destabilizing; as aggregation-prone if ΔAgg > 2, and otherwise non-aggregation-prone; and as hydrophobic if ΔHyd > 4, and otherwise non-hydrophobic else (Extended Data Fig. 4a, b). Solvent accessible surface area (ASA) for each residue was computed from the PDB structures with the DSSP program[76,77]. Residues were considered buried with ASA < 30, at the interface if the residue contributes ASA > 30 to the monomer interface in the assembled capsid, and as surface residue if ASA > 50 and not at the monomer interface.

To identify sites along the poliovirus sequence that differ significantly in their ability to accommodate destabilizing, aggregation-prone, or hydrophobic mutations, we next tested whether the mutation class E was statistically associated with the condition of either $Hsp90_N$ or $hsp90_i$ by constructing a 2×2 contingency table for each site composed of the mutation counts of E_$Hsp90_N$, NE_$Hsp90_N$, E_$hsp90_i$, and NE_$hsp90_i$. Finally, sites that are systematically enriched in destabilizing, aggregation-prone, or hydrophobic mutations across passages were identified by the CMH test on the $2 \times 2 \times 7$ contingency table per site and all seven sequenced passages. We found this the most robust approach considering experimental design, the nature of population sequencing and the possibility of spurious adaptation due to laboratory passaging[78]. After correcting p-values with the FDR method[70], significant sites are defined through $p < 0.05$ and counted $Hsp90_N$ if the odd's ratio (OR) > 1, and $hsp90_i$ for OR < 1. This identifies sites that are systematically, across passages, under selective pressure to differentially accommodate specific sets of mutations, namely destabilizing, aggregation-prone, or hydrophobic mutations as function of host-cell chaperone levels.

To identify individual mutations that give rise to significant sites as defined above, CMH test on the $2 \times 2 \times 7$ contingency tables were performed for every possible codon mutation along the poliovirus codon sequence comparing the mutation count and reference sequence codon count between $Hsp90_N$ and $hsp90_i$ as well as across passages. After p-value correction with the FDR method, we identified individual significant mutations through $p < 0.05$, as well as the requirement of them being at a significant site and following the same trend as the significant site, i.e., if a site is enriched in destabilizing mutations at $Hsp90_N$, then the individual mutations at the site also has to be destabilizing and enriched at $Hsp90_N$.

The mutational effect on codon optimality was tested based on the established tRNA adaptation index (tAI)[63] that quantifies how much coding sequences are adapted to the cellular tRNA pool for efficient translation. Here the tAI for HeLa

cells based on quantification of tRNA levels through tRNA microarrays[79] was used to compute a host-cell tAI. As a control, results were compared to those obtained with a tAI index computed from tRNA gene copy numbers in the human genome[80] (Supplementary Fig. 6). To assess global changes in codon optimality, we computed the $\Delta tAI = tAI_{mutation} - tAI_{reference}$ for each sequenced mutation in $Hsp90_N$ and $hsp90_i$ and analyzed the distributions of ΔtAI. To specifically test for an enrichment in mutations to more slowly translated non-optimal codons, we considered the bottom 20% codons with the lowest tAI values as non-optimal. The CMH test was used to identify sequence positions associated with mutations to non-optimal codons at $Hsp90_N$ and $hsp90_i$ across passages.

Code availability. Computer code to process raw sequencing data as well as reproduce the analysis has been deposited in the GitHub repositories https://github.com/ashleyacevedo/CirSeq and https://github.com/pechmann/polio.

Data availability. CirSeq data have been deposited in the Sequence Read Archive with accession code PRJNA438997. Other data are available from the corresponding authors upon request.

References

1. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. *Proc. Natl Acad. Sci. USA* **103**, 5869–5874 (2006).
2. DePristo, M. A., Weinreich, D. M. & Hartl, D. L. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat. Rev. Genet.* **6**, 678–687 (2005).
3. Zeldovich, K. B., Chen, P. & Shakhnovich, E. I. Protein stability imposes limits on organism complexity and speed of molecular evolution. *Proc. Natl Acad. Sci. USA* **104**, 16152–16157 (2007).
4. Gong, L. I., Suchard, M. A. & Bloom, J. D. Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* **2**, e00631 (2013).
5. Hartl, F. U., Bracher, A. & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324–332 (2011).
6. Tokuriki, N. & Tawfik, D. S. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* **459**, 668–673 (2009).
7. Aguilar-Rodríguez, J. et al. The molecular chaperone DnaK is a source of mutational robustness. *Genome Biol. Evol.* **8**, 2979–2991 (2016).
8. Bogumil, D. & Dagan, T. Cumulative impact of chaperone-mediated folding on genome evolution. *Biochemistry* **51**, 9941–9953 (2012).
9. Pechmann, S. & Frydman, J. Interplay between chaperones and protein disorder promotes the evolution of protein networks. *PLoS Comput. Biol.* **10**, e1003674 (2014).
10. Drake, J. W. The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes. *Ann. New Y. Acad. Sci.* **870**, 100–107 (1999).
11. Chen, P. & Shakhnovich, E. I. Lethal mutagenesis in viruses and bacteria. *Genetics* **183**, 639–650 (2009).
12. Domingo, E. & Holland, J. J. RNA virus mutations and fitness for survival. *Annu. Rev. Microbiol* **51**, 151–178 (1997).
13. Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E. & Andino, R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* **439**, 344–348 (2006).
14. Racaniello, V. R. in *Fields Virology* (eds Knipe, D. M. & Howley, P. M. eds) Ch. 24 (Lippincott Williams and Wilkins, Philadelphia, PA, 2013).
15. Schopf, F. H., Biebl, M. M. & Buchner, J. The HSP90 chaperone machinery. *Nat. Rev. Mol. Cell Biol.* **18**, 345–360 (2017).
16. Pearl, L. H. & Prodromou, C. Structure and mechanism of the Hsp90 molecular chaperone machinery. *Annu Rev. Biochem.* **75**, 271–294 (2006).
17. Taipale, M. et al. Quantitative analysis of Hsp90-client interactions reveals principles of substrate recognition. *Cell* **150**, 987–1001 (2012).
18. Geller, R., Vignuzzi, M., Andino, R. & Frydman, J. Evolutionary constraints on chaperone-mediated folding provide an antiviral approach refractory to development of drug resistance. *Genes Dev.* **21**, 195–205 (2007).
19. Neckers, L. Using natural product inhibitors to validate Hsp90 as a molecular target in cancer. *Curr. Top. Med Chem.* **6**, 1163–1171 (2006).
20. Yewdell, J. W. Viva la Revolución: rethinking influenza a virus antigenic drift. *Curr. Opin. Virol.* **1**, 177–183 (2011).
21. Martín, J., Crossland, G., Wood, D. J. & Minor, P. D. Characterization of formaldehyde-inactivated poliovirus preparations made from live-attenuated strains. *J. Gen. Virol.* **84**, 1781–1788 (2003).
22. Geller, R., Taguwa, S. & Frydman, J. Broad action of Hsp90 as a host chaperone required for viral replication. *Biochim Biophys. Acta* **1823**, 698–706 (2012).
23. Geller, R., Andino, R. & Frydman, J. Hsp90 inhibitors exhibit resistance-free antiviral activity against respiratory syncytial virus. *PLoS ONE* **8**, e56762 (2013).
24. Acevedo, A., Brodsky, L. & Andino, R. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* **505**, 686–690 (2013).
25. Williams, T. A. & Fares, M. A. The effect of chaperonin buffering on protein evolution. *Genome Biol. Evol.* **2**, 609–619 (2010).
26. Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
27. Queitsch, C., Sangster, T. A. & Lindquist, S. Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618–624 (2002).
28. Felix, M.-A. & Barkoulas, M. Pervasive robustness in biological systems. *Nat. Rev. Genet* **16**, 483–496 (2015).
29. Hummel, B. et al. The evolutionary capacitor HSP90 buffers the regulatory effects of mammalian endogenous retroviruses. *Nat. Struct. Mol. Biol.* **24**, 234–242 (2017).
30. Karras, G. I. et al. HSP90 shapes the consequences of human genetic variation. *Cell* **168**, 856–866 (2017).
31. Geiler-Samerotte, K. A., Zhu, Y. O., Goulet, B. E., Hall, D. W. & Siegal, M. L. Selection transforms the landscape of genetic variation interacting with Hsp90. *PLOS Biol.* **14**, e2000465 (2016).
32. Echave, J. & Wilke, C. O. Biophysical models of protein evolution: understanding the patterns of evolutionary sequence divergence. *Annu Rev. Biophys.* **46**, 85–103 (2017).
33. Serohijos, A. W. R. & Shakhnovich, E. I. Merging molecular mechanism and evolution: theory and computation at the interface of biophysics and evolutionary population genetics. *Curr. Opin. Struct. Biol.* **26**, 84–91 (2014).
34. Balch, W. E., Morimoto, R. I., Dillin, A. & Kelly, J. W. Adapting proteostasis for disease intervention. *Science* **319**, 916–919 (2008).
35. Knowles, T. P. J., Vendruscolo, M. & Dobson, C. M. The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* **15**, 384–396 (2014).
36. Hagan, M. F. & Chandler, D. Dynamic pathways for viral capsid assembly. *Biophys. J.* **91**, 42–54 (2006).
37. Jiang, P., Liu, Y., Ma, H.-C., Paul, A. V. & Wimmer, E. Picornavirus Morphogenesis. *Microbiol Mol. Biol. Rev.* **78**, 418–437 (2014).
38. Schymkowitz, J. et al. The FoldX web server: an online force field. *Nucleic Acids Res.* **33**, W382–W388 (2005).
39. Starr, T. N., Picton, L. K. & Thornton, J. W. Alternative evolutionary histories in the sequence space of an ancient protein. *Nature* **549**, 409–413 (2017).
40. Ashenberg, O., Gong, L. I. & Bloom, J. D. Mutational effects on stability are largely conserved during protein evolution. *Proc. Natl Acad. Sci. USA* **110**, 21071–21076 (2013).
41. Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J. & Serrano, L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* **22**, 1302–1306 (2004).
42. Szaruga, M. et al. Alzheimer's-causing mutations shift abeta length by destabilizing gamma-secretase-abetan interactions. *Cell* **170**, 443–456 (2017).
43. Gallardo, R. et al. De novo design of a biologically active amyloid. *Science* **354**, aah4949 (2016).
44. Dinner, A. R., Šali, A., Smith, L. J., Dobson, C. M. & Karplus, M. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem Sci.* **25**, 331–339 (2000).
45. Vassall, K. A. et al. Decreased stability and increased formation of soluble aggregates by immature superoxide dismutase do not account for disease severity in ALS. *Proc. Natl Acad. Sci. USA* **108**, 2210–2215 (2011).
46. Willmund, F. et al. The cotranslational function of ribosome-associated Hsp70 in eukaryotic protein homeostasis. *Cell* **152**, 196–209 (2013).
47. Olzscha, H. et al. Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. *Cell* **144**, 67–78 (2011).
48. Komar, A. A. A pause for thought along the co-translational folding pathway. *Trends Biochem Sci.* **34**, 16–24 (2009).
49. Zhang, G., Hubalewska, M. & Ignatova, Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.* **16**, 274–280 (2009).
50. Levy, E. D., De, S. & Teichmann, S. A. Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. *Proc. Natl Acad. Sci. USA* **109**, 20461–20466 (2012).
51. Pechmann, S., Willmund, F. & Frydman, J. The ribosome as a hub for protein quality control. *Mol. Cell* **49**, 411–421 (2013).
52. Nissley, D. A. & O'Brien, E. P. Timing is everything: unifying codon translation rates and nascent proteome behavior. *J. Am. Chem. Soc.* **136**, 17892–17898 (2014).

53. Chaney, J. L. & Clark, P. L. Roles for synonymous codon usage in protein biogenesis. *Annu. Rev. Biophys.* **44**, 143–166 (2015).

54. Pechmann, S. & Frydman, J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat. Struct. Mol. Biol.* **20**, 237–243 (2013).

55. Pechmann, S., Chartron, J. W. & Frydman, J. Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP in vivo. *Nat. Struct. Mol. Biol.* **21**, 1100–1105 (2014).

56. Tuller, T. et al. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**, 344–354 (2010).

57. Hussmann, J. A., Patchett, S., Johnson, A., Sawyer, S. & Press, W. H. Understanding biases in ribosome profiling experiments reveals signatures of translation dynamics in yeast. *PLoS Genet.* **11**, e1005732 (2015).

58. Dana, A. & Tuller, T. The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Res.* **42**, 9171–9181 (2014).

59. Sealfon, R. S. et al. FRESCo: finding regions of excess synonymous constraint in diverse viruses. *Genome Biol.* **16**, 38 (2015).

60. Jack, B. R. et al. Reduced protein expression in a virus attenuated by codon deoptimization. *G3* **7**, 2957–2968 (2017).

61. Lauring, AdamS., Acevedo, A., Cooper, SamanthaB. & Andino, R. Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus. *Cell Host Microbe* **12**, 623–632 (2012).

62. Han, Y. et al. Monitoring cotranslational protein folding in mammalian cells at codon resolution. *Proc. Natl Acad. Sci. USA* **109**, 12467–12472 (2012).

63. dos Reis, M., Savva, R. & Wernisch, L. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.* **32**, 5036–5044 (2004).

64. Aragones, L., Guix, S., Ribes, E., Bosch, A. & Pinto, R. M. Fine-tuning translation kinetics selection as the driving force of codon usage bias in the hepatitis A virus capsid. *PLoS Pathog.* **6**, e1000797 (2010).

65. Jahn, T. R. & Radford, S. E. Folding versus aggregation: polypeptide conformations on competing pathways. *Arch. Biochem. Biophys.* **469**, 100–117 (2008).

66. Tokuriki, N., Oldfield, C. J., Uversky, V. N., Berezovsky, I. N. & Tawfik, D. S. Do viral proteins possess unique biophysical features? *Trends Biochem. Sci.* **34**, 53–59 (2009).

67. Hartl, F. U. & Hayer-Hartl, M. Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* **16**, 574–581 (2009).

68. Taipale, M. et al. Chaperones as thermodynamic sensors of drug-target interactions reveal kinase inhibitor specificities in living cells. *Nat. Biotech.* **31**, 630–637 (2013).

69. Yam, A. Y. et al. Defining the TRiC/CCT interactome links chaperonin function to stabilization of newly made proteins with complex topologies. *Nat. Struct. Mol. Biol.* **15**, 1255–1262 (2008).

70. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).

71. Acevedo, A. & Andino, R. Library preparation for highly accurate population sequencing of RNA viruses. *Nat. Protoc.* **9**, 1760–1769 (2014).

72. Ribeiro, R. M. et al. Quantifying the diversification of hepatitis C virus (HCV) during primary infection: estimates of the in vivo mutation rate. *PLoS Pathog.* **8**, e1002881 (2012).

73. Nei, M. & Gojobori, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**, 418–426 (1986).

74. Kryazhimskiy, S. & Plotkin, J. B. The population genetics of dN/dS. *PLoS Genet.* **4**, e1000304 (2008).

75. Pechmann, S., Levy, E. D., Tartaglia, G.-G. & Vendruscolo, M. Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. *Proc. Natl Acad. Sci. USA* **106**, 10159–10164 (2009).

76. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).

77. Joosten, R. P. et al. A series of PDB related databases for everyday needs. *Nucleic Acids Res.* **39**, D411–D419 (2011).

78. McWhite, C. D., Meyer, A. G. & Wilke, C. O. Sequence amplification via cell passaging creates spurious signals of positive adaptation in influenza virus H3N2 hemagglutinin. *Virus Evol.* **2**, vew026 (2016).

79. Pavon-Eternod, M. et al. tRNA over-expression in breast cancer and functional consequences. *Nucleic Acids Res.* **37**, 7268–7280 (2009).

80. Chan, P. P. & Lowe, T. M. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* **37**, D93–D97 (2009).

Acknowledgements

Author contributions

R.G., J.F., and R.A. designed the study; R.G., S.P., J.F., and R.A. interpreted the data and wrote the manuscript. R.G. carried out all experiments and S.P. the computational analyses; A.A. performed the CirSeq sequencing; R.A. and J.F. contributed equally to the study.

Additional information