Continuous directed evolution of proteins with improved soluble expression

Tina Wang[1,2], Ahmed H. Badran[1,2], Tony P. Huang[1,2] and David R. Liu[1,2,3]*

We report the development of soluble expression phage-assisted continuous evolution (SE-PACE), a system for rapidly evolving proteins with increased soluble expression. Through use of a PACE-compatible AND gate that uses a split-intein pIII, SE-PACE enables two simultaneous positive selections to evolve proteins with improved expression while maintaining their desired activities. In as little as three days, SE-PACE evolved several antibody fragments with >5-fold improvement in expression yield while retaining binding activity. We also developed an activity-independent form of SE-PACE to correct folding-defective variants of maltose-binding protein (MBP) and to evolve variants of the eukaryotic cytidine deaminase APOBEC1 with improved expression properties. These evolved APOBEC1 variants were found to improve the expression and apparent activity of Cas9-derived base editors when used in place of the wild-type cytidine deaminase. Together, these results suggest that SE-PACE can be applied to a wide variety of proteins to rapidly improve their soluble expression.

Soluble protein expression is a critical requirement for the production and application of proteins. *Escherichia coli* is generally the most convenient and commonly used organism for protein expression. It is estimated, however, that <50% of bacterial and <15% of nonbacterial proteins can be expressed in soluble form in *E. coli*; common barriers to soluble expression include proteolysis or misfolding into inclusion bodies[1]. Moreover, engineering or evolving proteins toward improved or novel functions can often lead to reduced soluble expression, impeding the development and application of proteins with tailor-made functional properties[2]. Though it is possible to optimize expression conditions[3], vary host strains[4], or fuse the protein of interest (POI) to solubility-enhancing tags[5], an attractive complementary approach is to engineer or evolve the POI itself to improve its soluble expression under conditions that preserve desirable preexisting activities.

To this end, researchers have successfully used directed evolution to identify protein variants that are resistant to high temperatures[6], denaturants[7], or proteases[8]. The resulting proteins possess enhanced thermodynamic stability, a property that often, but not always, leads to improved soluble expression[2]. Screening protein libraries in whole cells or cell lysates to identify highly active variants can also coincidentally yield proteins with superior expression[9], though this approach is often confounded by variants with higher activity and no improvements in expression. Finally, folding reporters measure soluble protein expression levels through fusion of the POI to a reporter protein such as GFP[10,11]. These reporters enable the selection of soluble variants outside of the context of protein activity, offering broad applicability but with potential loss of function.

In addition to the challenge associated with simultaneously yet independently selecting for soluble expression and protein function, traditional directed evolution approaches to improving protein expression introduce additional drawbacks, including substantial time and effort requirements[12]. Each round of traditional laboratory protein evolution can take a week or longer, rendering the prospect of evolving improvements in soluble expression in addition to

new or improved functions an unattractive one. As a result, otherwise interesting targets for protein evolution are often not pursued because of the intractability of expression under standard laboratory conditions. A method that allows the rapid improvement of protein expression coupled with the preservation or improvement of protein function would therefore offer substantial benefits.

Here we report a phage-assisted continuous evolution (PACE) system for rapidly evolving proteins that have improved soluble protein expression in *E. coli*, either in the presence or absence of a simultaneous selection for protein function. The system uses an AND logic gate that enables PACE under two concurrent positive selections. We applied this system toward the evolution of antibody single-chain variable fragments (scFv) that canonically misfold and aggregate in the *E. coli* cytoplasm. In as little as 3 d, our system evolved several scFv variants with improved cytosolic expression and little to no loss of binding activity. Additionally, using disulfide-free scFvs, our PACE system evolved variants with enhanced thermodynamic stability in as little as 5 d. In an activity-independent mode, the system yielded APOBEC1 cytidine deaminase variants with improved solubility that enhanced the purification yields and editing activity of base editors in *E. coli* and mammalian cells. Together, these results establish SE-PACE as a rapid method to improve the expression and, in some cases, the stability of a variety of proteins while preserving their function.

Results

A split T7 RNAP reporter for soluble expression PACE. In PACE (Fig. 1a), a population of filamentous bacteriophage (selection phage, SP) is continuously diluted out of a fixed-volume vessel (the lagoon) by host *E. coli* cells. Engineered SPs carry the evolving gene of interest in place of gene III (gIII), which encodes the minor coat protein III (pIII) that is required to generate infectious progeny phage. SPs containing desired (active) variants of the gene of interest trigger gIII expression and pIII production from an accessory plasmid (AP) in the host cells. Progeny phage production scales with pIII levels[13]. As a result, SPs encoding desired target gene
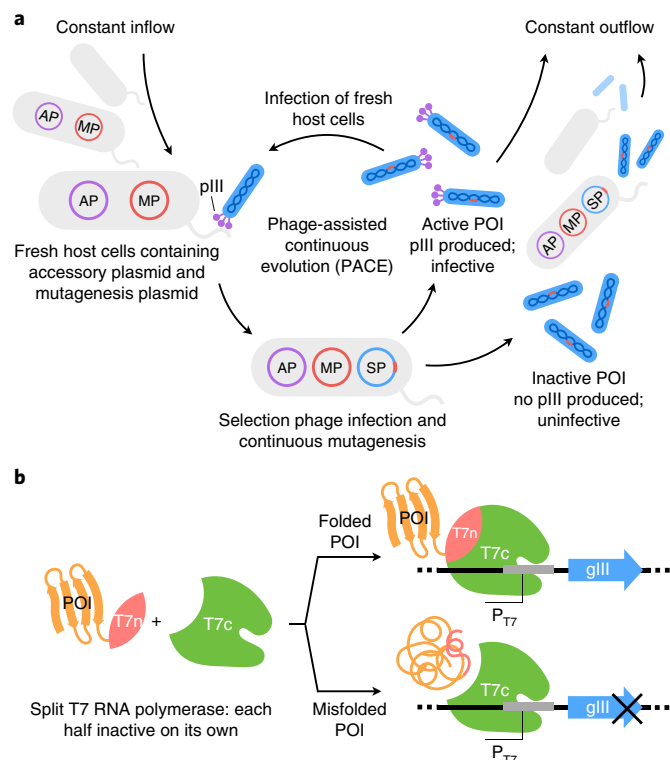
Fig. 1 | Overview of a PACE selection for improved soluble protein expression. a, General PACE schematic. Host cells contain an accessory plasmid (AP) that expresses gene III (gIII) in the presence of a desired protein of interest (POI). The selection phage (SP) contains the POI in place of gIII and can therefore only produce infectious progeny if it encodes a POI variant that passes the selection established by the AP, thereby triggering pIII production. The gene encoding the POI is mutated by induction of the mutagenesis plasmid (MP). As the SP exist in a continuously diluted fixed-volume vessel (the lagoon), only those phage genotypes that propagate faster than the rate of dilution can persist in the lagoon. **b**, A folding reporter approach links soluble protein expression to production of protein pIII, required for phage infection. Soluble expression of the POI causes persistence of the fused N-terminal half of T7 RNA polymerase (POI–T7n). A misfolded, insoluble, or proteolyzed POI produces nonfunctional or degraded T7n, which is unable to associate with the C-terminal half of T7 RNAP (T7c) and thus unable to transcribe gIII.

variants produce more infectious offspring capable of replicating faster than they are diluted from the lagoon. Diversity is generated through induction of a mutagenesis plasmid (MP) that dramatically increases SP mutation rates[14]. PACE has been successfully applied to a wide range of proteins, including polymerases[15–17], proteases[18,19], genome-editing proteins[20], insecticidal toxins[21], and aminoacyl-tRNA synthetases[22].

To link target protein expression to the production of pIII and phage propagation, we sought to render T7 RNA polymerase (T7 RNAP) activity dependent on the soluble expression of a target protein (Fig. 1b) and use the resulting activity to drive gene III transcription[15–19,22]. Although full-length T7 RNAP (883 amino acids) is too large to use as a folding reporter[11], T7 RNAP can be split between amino acids 179 and 180 to produce two fragments that are individually inactive but spontaneously associate into functional T7 RNAP[23,24].

We reasoned that by fusing the smaller N-terminal fragment of this split T7 RNAP (T7n, residues 1–179) to the C-terminus of the protein of interest (POI) and providing the larger T7 RNAP

fragment (T7c, residues 180–883) in the host cell, expression of a P_{T7}-gIII transcript would be dependent on soluble POI expression. Previous folding reporter studies[11] suggest that the overall folded state of the POI–T7n fusion should be strongly determined by the degree to which the N-terminal POI is soluble and well-folded. Therefore, SP encoding a soluble, well-folded POI–T7n fusion should allow the T7n domain to associate with T7c, restoring T7 RNAP-mediated transcription of gIII from its T7 promoter (P_{T7}; Fig. 1b) and enabling propagation of SP encoding the POI during PACE (Fig. 1a). In contrast, SP encoding POIs that express poorly or are unstable or prone to aggregation or proteolysis should result in less available POI–T7n and reduced T7 RNAP activity and thus fewer progeny phage (Fig. 1a,b).

To test this SE-PACE selection system, we performed transcriptional activation assays in which bacterial luciferase (rather than gene III) expression is controlled by P_{T7} (Supplementary Fig. 1). We fused a variety of known MBP mutants with variable soluble expression properties[25,26] to T7n. We observed a correlation between luminescence signal and soluble expression levels of the MBP mutants (Fig. 2a and Supplementary Fig. 2a), as well as a weaker relationship between luminescence and the known thermodynamic stability of the MBP mutants (Fig. 2b). These results demonstrate that POI–T7n fusions allow split T7 RNAP activity to be linked primarily to the soluble expression levels and, to a lesser extent, the thermodynamic stability of a model POI.

Next, we performed competitive phage propagation experiments using SP encoding wild-type MBP or a destabilized mutant (G32D + I33P) in a PACE format. Unlike wild-type MBP, the G32D + I33P mutant has been shown to express primarily in inclusion bodies[25]. SP encoding wild-type MBP-T7n selectively propagated at the expense of SP encoding MBP(G32D + I33P)-T7n on host *E. coli* cells containing an AP that constitutively expresses T7c and provides gIII from P_{T7}. After 45 h of continuous propagation at a flow rate of 1 lagoon vol/h, corresponding to 23 average generations of phage reproduction, the SP encoding wild-type MBP were enriched at least 10,000-fold over the SP encoding the mutant MBP (Supplementary Fig. 2b). These findings establish the ability of SE-PACE to support the selective propagation of SPs encoding soluble proteins in a PACE format.

Finally, we applied this selection to evolve variants with improved solubility, starting with the two most poorly expressed MBP mutants tested, Y283D and G32D + I33P, in PACE (Fig. 2c) using enhanced mutagenesis (mutagenesis plasmid MP6, ref. [14]) and a flow rate of 0.5–2 lagoon vol/h. After 72 h (80 average generations) of PACE, we observed complete reversion of the Y283D mutant to wild-type MBP (Supplementary Fig. 3a). The lagoon seeded with the MBP(G32D + I33P) mutant evolved a P33T substitution at the critical position 33, which improved expression levels by ~2-fold (Supplementary Fig. 3b). Together, these results suggest that the split T7 RNAP folding reporter can be applied to evolve model proteins with improved soluble expression in PACE.

Development of an AND gate for dual selection in PACE.
Attempts to further evolve the MBP(G32D + I33T) mutant through SE-PACE with increased selection stringency, achieved through use of APs encoding deactivating mutations in T7c (Supplementary Fig. 4a), predominantly led to the evolution of frameshift mutations in the latter half of MBP. These frameshifts cause premature stop codons upstream of ATG (Met) codons present in the native protein sequence, enabling translational re-initiation at these sites to produce N-terminally truncated T7n fusions that allow the mutant SP to survive the selection in a manner that is not dependent on POI expression (Supplementary Fig. 4b). To overcome this challenge, we sought to add a simultaneous selection for POI function, as truncation products would likely experience defects in activity. Selecting both for soluble expression of T7n and for POI function
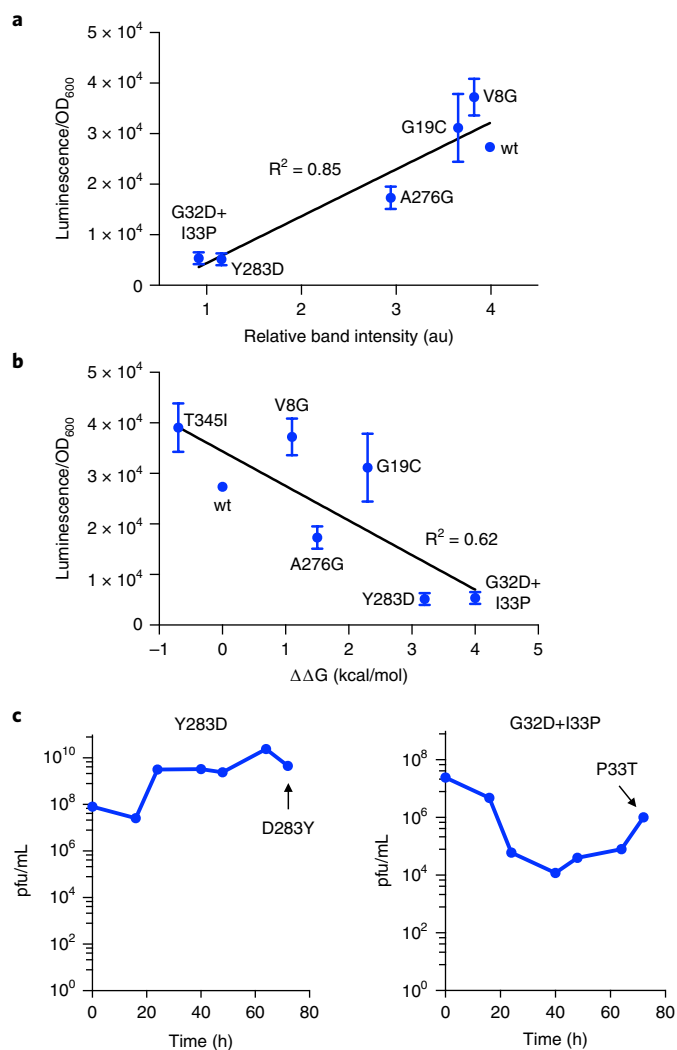
Fig. 2 | Validation of the soluble protein expression PACE selection.
a, Soluble expression of MBP variants in *E. coli* at 37 °C, as revealed by SDS–PAGE analysis of soluble protein extracts (Supplementary Fig. 2a), correlates with transcriptional activity ($R^2 = 0.85$). Luminescence data reflects the mean and s.d. of three technical replicates (unique clones). **b**, Transcriptional activity is also loosely correlated with thermodynamic stability of MBP variants ($R^2 = 0.62$). Luminescence data reflects the mean and s.d. of three technical replicates (unique clones). **c**, SP encoding the poorly expressed D283Y and G32D + I33P MBP mutants evolve expression-enhancing mutations in PACE over 3 d. The *y* axis shows total phage titer in each (*N* = 1) lagoon.

should also reduce the likelihood of losing protein function while improving expression, a known disadvantage of activity-independent folding reporters[27].

To perform two simultaneous positive selections during PACE, we used a trans-splicing split intein to create an AND gate for pIII-dependent phage propagation. Although a split intein T7 RNAP has been previously reported[28], we reasoned that an AND gate constructed using pIII itself would be compatible with a wider scope of PACE selections. To implement this strategy, we divided pIII into two fragments, each fused to one half of the split intein. Each PACE positive selection leads to the production of one of the two split-intein–pIII halves. If the POI passes only one selection, the other split-intein–pIII half will not be produced, and trans-splicing to generate pIII therefore cannot take place. In contrast, a POI capable

of passing both selections results in production of both halves of the split-intein–pIII, allowing the trans-splicing reaction to generate full-length pIII necessary for progeny phage production (Fig. 3a).

Because pIII is normally directed to the periplasm after translation[29], we chose to split pIII within its signal peptide sequence to generate translocation-competent pIII only upon trans-splicing. To determine the optimal site, we inserted a Cys-Phe-Asn sequence (CFN; the native C-extein scar left behind by the fast-splicing cyanobacteria DnaE split inteins[30]) at various positions in the hydrophobic region of the signal peptide and evaluated its effect on phage propagation. Insertion of CFN after Ile8 or Leu10 was well-tolerated (Supplementary Fig. 5), suggesting that pIII might be dividable at these positions. Next, we split pIII between amino acids 10–11 of its signal peptide and fused each half to the *Nostoc punctiforme* (Npu) DnaE split intein[31] to create split-intein–pIII halves. We expressed both halves of split-intein–pIII from the phage shock protein promoter (P_{PSP})[15,32] on two separate APs (AP1 and AP2) and tested whether both APs in combination could support the propagation of SP encoding a kanamycin resistance gene in place of gIII (Fig. 3b). Host cells co-transformed with both AP1 and AP2 supported robust phage propagation at a level comparable to that in host cells transformed with a single AP containing full-length gIII. Providing only one half of the split-intein pIII (AP1 only or AP2 only), or mutating residues critical to trans-splicing, resulted in negligible phage propagation, demonstrating the AND gate-like behavior of the split-intein–pIII system (Fig. 3b).

Next, we tested whether the split-intein–pIII could link two different selections with two separate APs, each expressing one half of split-intein–pIII under control of a selection for protein binding activity or protein expression (Fig. 3c). For protein binding activity, we used the HA4 monobody–ABL1 kinase SH2 domain interaction[33], as we previously showed that SP encoding HA4 propagate robustly on an AP containing a 434cI–SH2 domain fusion[21]. Because monobodies are typically highly soluble, we expected HA4 to also pass the protein expression selection. Indeed, only SP encoding a T7n–HA4 fusion, which passes both the protein binding and soluble expression selections, form plaques on cells harboring the two split-intein–pIII APs (Fig. 3c). Conversely, SPs that pass only one selection (encoding either HA4 or T7n alone) exhibit no plaque activity. These results suggest that the split-intein–pIII system should enable two distinct, simultaneous positive selections in PACE. Finally, we used split-intein–pIII to evolve the binding-impaired HA4 Y87A mutant[33] and successfully rescued its ABL1 SH2 domain binding activity (Supplementary Fig. 6), demonstrating the ability of our system to restore targeted protein binding activity in PACE.

PACE with the split-intein–pIII dual-selection system. scFvs are fusion proteins that consist of the V_H and V_L variable domains of antibodies joined with a flexible linker[34]. Although they retain the epitope binding affinity and specificity of their parent antibodies, scFvs offer several advantages including substantially smaller size and compatibility with expression in *E. coli*. Typically, the V_H and V_L domains each contain a stabilizing disulfide bond unable to form in the reducing environment of the *E. coli* cytosol, leading to very poor expression. Because of this limitation, engineering and evolving these proteins to improve their cytosolic expression has been the focus of major effort[35].

We hypothesized that scFv soluble expression could be improved while preserving binding activity using the dual-selection PACE system (Fig. 4a). We began with an scFv targeting the yeast GCN4 transcription factor leucine zipper[36]. Initially isolated through ribosome display, this scFv was extensively engineered for improved expression to ultimately yield the Ω graft (Ωg) variant[37]. Despite these efforts, the majority of Ωg remains insoluble in the *E. coli* cytosol[38]. We sought to improve the soluble expression of this scFv with SE-PACE.
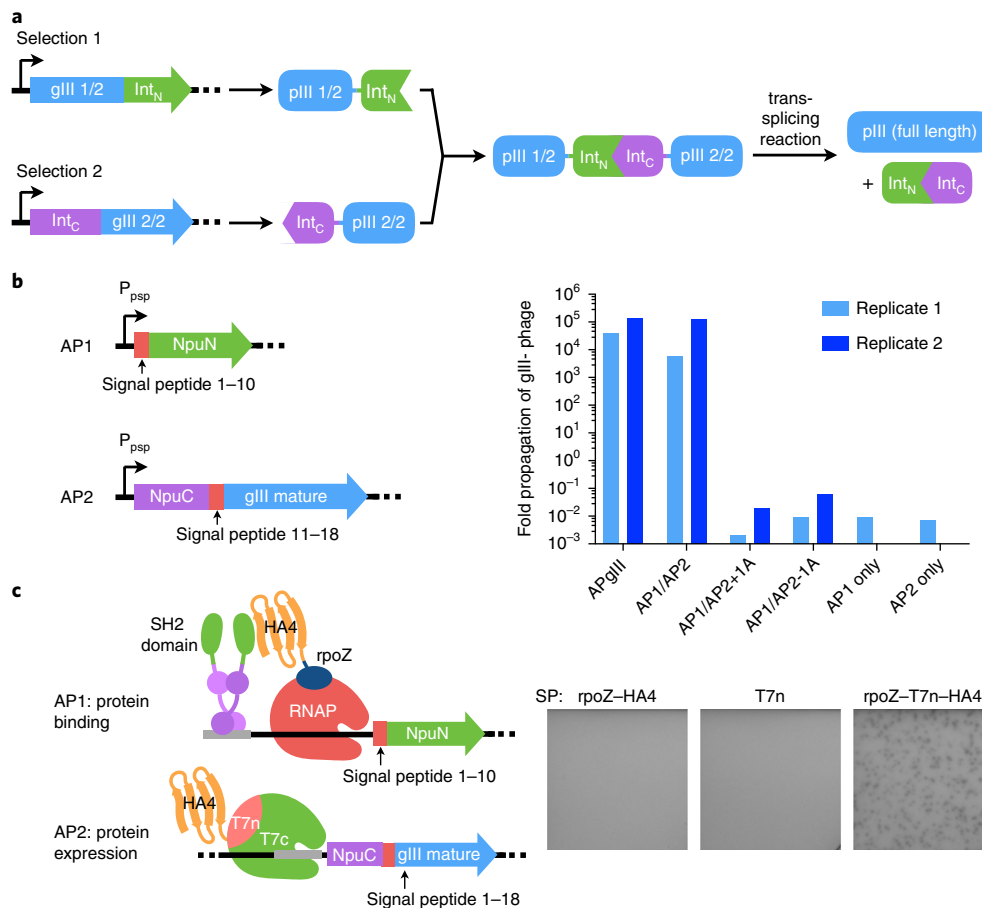
Fig. 3 | Split-intein pIII as an AND gate for bridging two positive selections in PACE. a, Each positive selection produces one half of split-intein pIII, which can only undergo trans-splicing and produce functional pIII when both halves are present, thus requiring activity on both selections to generate infectious progeny phage. **b**, Split-intein pIII supports robust phage propagation in a trans-splicing-dependent manner. Negligible propagation is observed when only one half is present (AP1 only or AP2 only), or when residues critical to splicing are mutated (AP1/AP2$_{+1A}$ or AP1/AP2$_{-1A}$). Two biological replicates are shown (enrichments performed on two separate days). **c**, Split-intein pIII acts as an AND gate between two orthogonal PACE positive selections. The canonical gIII sequence is replaced by either half of the split-intein gIII in APs for protein binding (AP1) and protein expression (AP2). Only SP that encode proteins passing both selections (rpoZ-T7n-HA4) produce infectious progeny, resulting in plaque formation on host cells containing AP1 and AP2, while SP that pass only a single selection (rpoZ-HA4 or T7n) do not form plaques. Scale bar, 10 mm. This experiment was not repeated.

We evolved SP encoding Ωg in SE-PACE, selecting for both protein-binding activity and expression. The protein-binding selection used a GCN4 leucine zipper peptide fused to 434cI, whereas the SP carried Ωg fused to rpoZ (RNA polymerase omega subunit) as well as a PACE-optimized version of T7n (eT7n; Supplementary Fig. 7) to minimize fitness gains that could be evolved in T7n independent of mutations in the POI. After 72 h (31 average generations), we observed enrichment of phage carrying several mutations in Ωg enhancing protein expression to a level comparable to that of m3, a recently identified Ωg variant with greatly improved cytosolic expression (Fig. 4b and Supplementary Fig. 8a,b). The best PACE-evolved clones, 29.1.2 (S38L + V40I + F87S + N190D) and 29.1.5 (F87S + G103S + L224V) improved expression > 2.5-fold, as observed by gel densitometry (Supplementary Fig. 9a–c) and increased expression yields up to 5-fold (with a range of 2- to 17-fold) (Supplementary Table 1). The individual mutations comprising 29.1.5 were found to have a positive effect on expression levels when introduced into Ωg; however, none of these single point mutants expressed as well as the triple F87S + G103S + L224V mutant (Supplementary Fig. 9f,g and Supplementary Table 1). The binding affinities of Ωg and 29.1.5 for the GCN4 leucine zipper sequence were comparable by

ELISA, whereas 29.1.2 exhibited approximately five-fold lower affinity (Fig. 4c and Supplementary Table 1).

Our PACE-evolved variants showed slightly decreased (~2 °C lower) melting temperatures (T_m) compared to Ωg (Supplementary Table 1 and Supplementary Fig. 9e). This may result from the discrepancy between conducting T_m measurements on proteins evolving in an environment that requires folding without disulfide formation and on purified protein under oxidizing conditions. To probe the disulfide dependence of scFvs evolved in SE-PACE, we mutated the disulfide-forming cysteines in m3 and 29.1.5 to serines (denoted m3-noCys and 29.1.5-noCys, respectively), decreasing the T_m of each scFv by > 10 °C (Supplementary Table 1). PACE of these disulfide-free scFvs selecting for both protein binding and expression resulted in variants with improved thermodynamic properties (Supplementary Fig. 10 and Supplementary Table 1). For 29.1.5-noCys, this effect was dramatic, with one clone (58.3.1) possessing an improvement in T_m of > 10 °C. Additionally, variants 58.1.1 and 58.1.2 showed a + 5 °C increase in T_m when compared to m3-noCys (Supplementary Table 1). These results suggest that, despite selecting primarily for protein expression, our system is capable of producing variants with improved thermodynamic stability when other factors influencing stability such as disulfide bond formation are removed.
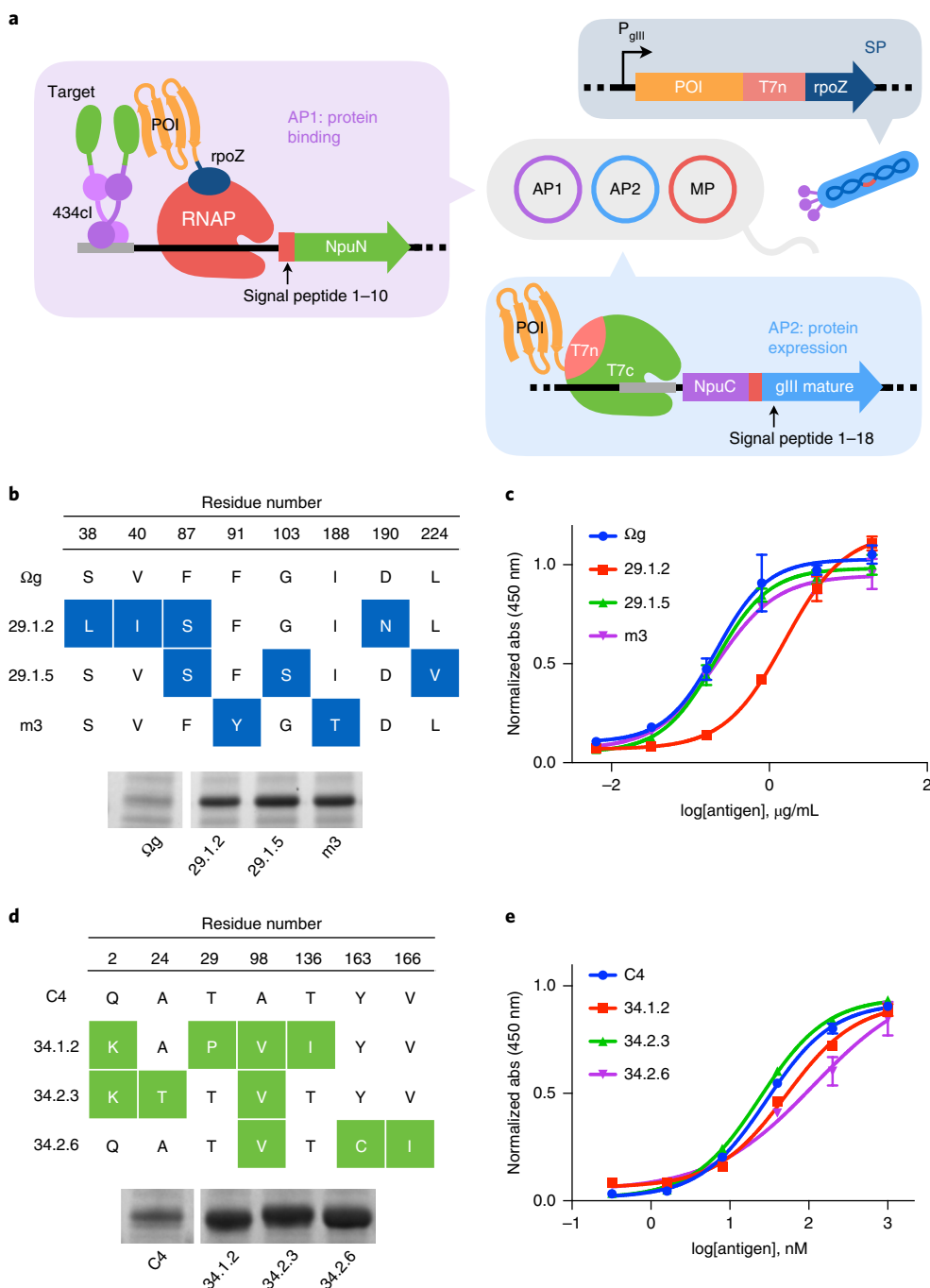
Fig. 4 | Soluble expression PACE with simultaneous selection for expression and binding activity. a, Overview of SE-PACE with two positive selections. In this case, a selection for POI protein binding activity (AP1) is run simultaneously with a selection for POI protein expression (AP2). **b**, PACE of the anti-GCN4 leucine zipper scFv Ωg produces variants 29.1.2 and 29.1.5 (mutations highlighted in blue), which show increased soluble expression at 37 °C. For comparison, m3 is a previously reported mutant[38] with greatly improved soluble expression. The full gel is provided in Supplementary Fig. 9a. **c**, Evolved variants maintain binding affinity for the GCN4 leucine zipper peptide motif. Data reflects the mean and s.d. of three technical replicates (three separate ELISA wells using protein from the same preparation). This experiment was repeated once with similar results. **d**, PACE of the anti-Htt scFv C4 results in mutations (highlighted in green) that greatly enhance soluble expression at 37 °C. The full gel is provided in Supplementary Fig. 12a. **e**, Binding of PACE-evolved variants and C4 to a peptide corresponding to residues 1–17 of Htt exon 1. Data reflects the mean and s.d. of three technical replicates (three separate ELISA wells using protein from the same preparation). This experiment was repeated once with similar results.

Finally, we used the dual PACE selection system to evolve a second scFv, C4, which recognizes the first 17 amino acids of the huntington (Htt) protein and has been shown to reduce aggregation of Htt fragments with large polyglutamine expansions[39]. One hundred and twenty hours (70 average generations) of dual selection PACE with a protein-binding AP encoding the first 17 amino acids

of Htt and a moderately stringent soluble expression AP enriched for two different mutations that each increased soluble expression (Supplementary Figs. 11a and 12a). An additional 114 h (94 average generations) of PACE using the same protein-binding AP and a more stringent protein expression AP produced genotypes that further increased soluble expression by > 2-fold compared to
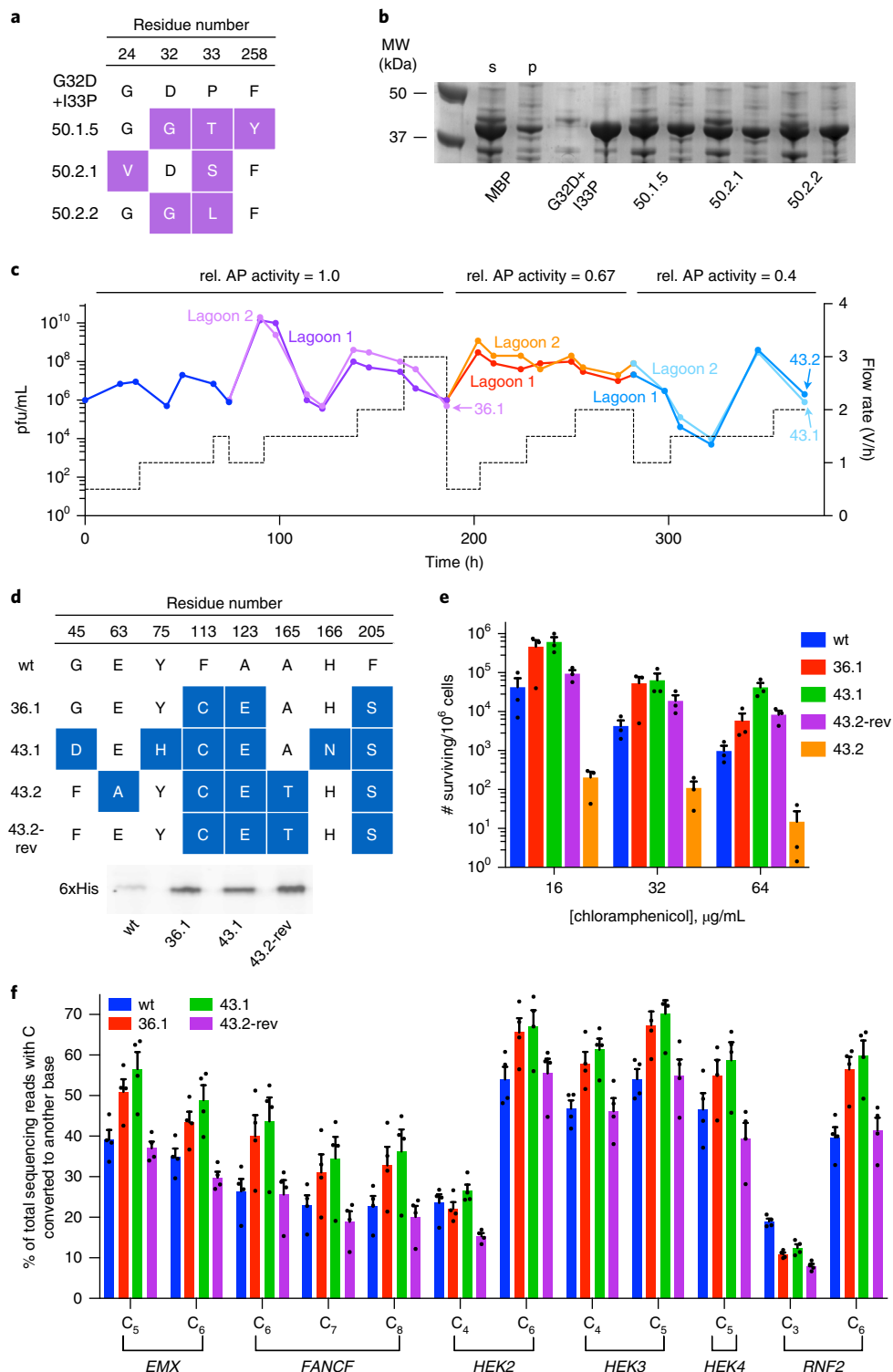
Fig. 5 | Activity-independent selection for soluble proteins in PACE. a, SE-PACE of the MBP(G32D + I33P) mutant produces partial reversions to the wild-type sequence after 4 d. Mutations are highlighted in purple. **b**, PACE-evolved MBP variants show greatly improved soluble expression in *E. coli* at 37 °C when compared to starting MBP(G32D + I33P). The full gel is provided in Supplementary Fig. 15a. "s" = soluble fraction, "p" = pellet. This experiment was not repeated. **c**, PACE of cytidine deaminase rAPOBEC1. The stringency of the protein expression selection was modulated through increasing the rate of lagoon dilution (flow rate; dashed lines) as well as the use of activity-diminishing T7 RNAP and T7 promoter mutants in the AP. **d**, rAPOBEC1 variants obtained after 186 h and 370 h (mutations highlighted in blue) show greatly improved soluble expression at 37 °C when compared to wild-type. The full blot is provided in Supplementary Fig. 16a. This expression experiment was repeated once with similar results. **e**, Base editor (BE2) variants employing evolved rAPOBEC1s show enhanced apparent activity in *E. coli*, as measured by cells surviving selection on increasing concentrations of chloramphenicol due to BE2-dependent reversion of an active site mutation in chloramphenicol acetyl transferase. Data reflects the mean and s.d. of three technical replicates (unique clones). **f**, Base editor (BE3) variants using evolved rAPOBEC1 36.1 and 43.1 variants show enhanced editing efficiency in HEK293T cells, as measured by the percentage of total DNA sequencing reads with Cs in the indicated target positions converted to another base. The sequences of the six genomic loci tested (*EMX–RNF2*) are provided in Supplementary Fig. 19a. Data reflects the mean and s.e.m. of four biological replicates (experiments performed on different days).

C4 by gel densitometry (Fig. 4d and Supplementary Fig. 12a) and increased expression yields up to 6-fold (Supplementary Table 2). Similar to the PACE-evolved Ωg variants, these mutants did not possess higher stability (Supplementary Table 2). Two evolved anti-Htt scFvs (34.1.2 and 34.2.3) exhibited similar binding to that of C4 to a peptide containing the first 17 amino acids of Htt, whereas the third characterized variant (34.2.6) showed slightly reduced affinity (Fig. 4e and Supplementary Table 2). Together, these results suggest that dual selection SE-PACE is capable of producing improvements in protein expression with little or no loss of binding activity.

Effect of activity selection on expression improvements. In some directed evolution experiments[9], including ones employing PACE[19], an activity-based selection alone can be sufficient to produce gains in expression. To investigate the contribution of the activity selection to the improvement of protein expression, we evolved MBP(G32D + I33T) for binding to the anti-MBP monobody YSX1 (ref. [40]). We performed two parallel PACE experiments: one simultaneously selecting for protein binding and protein expression (P26) and one selecting for protein binding alone (using a protein expression AP with zero stringency, P28). Most of the mutations fixed in both experiments mapped to the YSX1–MBP binding interface (Supplementary Fig. 13a,b), suggesting that strong selection pressure for protein binding was present. However, the phage pool that experienced selection for protein expression enriched an additional mutation at Gly32 that was not present in the phage pool selected with protein binding alone (Supplementary Fig. 13a). Variants from P28 exhibited improved soluble expression in comparison to MBP(G32D + I33T) (Supplementary Fig. 13c). However, variants from P26 containing the G32D mutation showed even greater increases in expression (Supplementary Fig. 13c); in the case of 26.5, soluble expression was comparable to that of wild-type MBP. These results confirm previous observations[2] that selections for activity can increase soluble protein expression, but also suggest that a dedicated selection for soluble expression can further improve this property.

Activity-independent soluble expression PACE. Although the simultaneous positive selection approach possesses the advantage of preserving protein activity while improving expression, it requires an additional PACE selection for activity. An activity-independent selection for improving protein expression might provide a useful complement to the dual-selection system. To achieve a useful activity-independent SE-PACE selection, we sought to address the primary source of 'cheating' observed with the split T7 RNAP folding reporter—the formation of premature truncation products upstream of translation-initiating Met residues. We hypothesized that this problem could be circumvented through judicious application of the protein-binding selection, which requires the translation of rpoZ (the omega subunit of *E. coli* RNA polymerase) fused to the target protein. If we added a small affinity tag at the N terminus of the target protein in the POI–T7n fusion and fused rpoZ to the C terminus, then the resulting tag–POI–T7n–rpoZ construct must be translated in its entirety to maintain covalent linkage of the tag and rpoZ. In this system, truncations in the middle of the POI–T7n fusion would separate the tag and rpoZ, impairing the ability of the POI to recruit rpoZ to a tag-binding protein localized to the gIII promoter (Supplementary Fig. 14).

We used the GCN4 leucine zipper peptide epitope as our affinity tag and the anti-GCN4 scFv clone m3 as the tag-binding protein fused to 434cI, thereby localizing tag-linked proteins to the gIII promoter (Supplementary Fig. 14). To validate this activity-independent, cheater-resistant selection strategy, we returned to the evolution of the G32D + I33P mutant of MBP, which was originally frustrated by enrichment of truncating genotypes. An SP encoding the GCN4 peptide–[MBP(G32D + I33P)]–eT7n–rpoZ quadruple

fusion was subjected to PACE, with host cells transformed with split-intein pIII APs selecting for both GCN4 peptide binding and soluble protein expression. After 96 h, we obtained enrichment of SPs bearing mutations fully reverting Asp32 and substituting either Thr (50.1.5) or Leu (50.2.2) for Pro33 (Fig. 5a). A third genotype, G24V + P33S (50.2.1), was also observed. All three genotypes showed greatly enhanced (>10-fold) expression relative to the starting G32D + I33P mutant (Fig. 5b and Supplementary Fig. 15a). 50.1.5 and 50.2.2 also possessed higher thermodynamic stability than G32D + I33P (Supplementary Table 3). We did not observe full reversion to the wild-type genotype, likely because a Pro to Ile substitution requires changing three adjacent nucleotides, whereas substitution to Thr or Leu requires mutation of only a single nucleotide and is more frequently accessed during random mutagenesis. Notably, no truncating genotypes were observed among SP surviving this PACE, suggesting that the peptide affinity tag strategy enabled successful evolution of protein expression in the absence of a selection for protein activity without evolving cheaters.

Finally, we applied this cheater-resistant activity-independent protein expression evolution to improve the soluble expression of rat apolipoprotein B mRNA editing catalytic subunit 1 (rAPOBEC1). APOBEC1 is a potent cytidine deaminase that can act on both RNA and DNA. Like many eukaryotic proteins, rAPOBEC1 expresses poorly in *E. coli* and localizes almost exclusively to the insoluble fraction (Supplementary Fig. 16a). We evolved SP encoding rAPOBEC1 for a total of 370 h (260 average generations) in PACE sequentially using three protein expression APs of increasing stringency (Fig. 5c). After the first 74 h, SPs containing the two dominant genotypes were pooled, and this mixture was used to infect two parallel, identically run lagoons. At 186 h, both lagoons converged on mutations (F113C + A123E + F205S) that yielded the 36.1 genotype (Fig. 5d); however, lagoon 1 had also collected an inactivating mutation (E63A). The final 184 h of PACE produced two variants (43.1 and 43.2) that enhanced expression >4-fold (Fig. 5d and Supplementary Fig. 16a). As 43.2 was identified from lagoon 1, it also carried the E63A mutation, which was reverted to obtain 43.2-rev (Fig. 5d). When tested in a rifampicin resistance assay[41], neither 36.1 nor 43.1 showed defects in apparent deaminase activity. Evolved APOBEC clone 43.2-rev also showed comparable activity to wild-type rAPOBEC1, whereas 43.2 was completely inactive, as expected (Supplementary Fig. 17a). Clones 36.1, 43.1, and 43.2-rev all had a negative impact on cell growth rate when expressed, whereas wild-type rAPOBEC1 and 43.2 exhibited no such effect, suggesting that the observed growth inhibition may be due to increased expression of active deaminase (Supplementary Fig. 17b).

These SE-PACE-evolved APOBEC1 variants also increased soluble expression levels when incorporated into base editors, engineered genome-editing proteins consisting of a fusion of rAPOBEC1, a catalytically impaired Cas9, and a uracil glycosylase inhibitor that enable targeted conversion of individual DNA base pairs in living cells[42–45]. Substitution of wild-type rAPOBEC1 with 36.1, 43.1, and 43.2-rev in the base editor BE3 improved soluble expression yields by 2–3-fold (Supplementary Table 4 and Supplementary Fig. 18), despite the fact that rAPOBEC1 only accounts for ~15% of the total BE3 protein by molecular weight. Finally, base editors incorporating evolved APOBEC variants 36.1, 43.1, and 43.2-rev exhibited higher apparent editing activity in *E. coli*, as measured by the ability of BE2 (ref. [42]), which contains dCas9 in place of Cas9 nickase, to rescue an active site mutation (H193R) in chloramphenicol acetyl transferase[44] (Fig. 5e). Increased cytidine deamination was also observed in HEK293T cells transfected with BE3 variants 36.1 and 43.1 (Fig. 5f), albeit with modestly decreased product purities (Supplementary Fig. 19a). Because these APOBECs were not selected for deaminase activity during PACE, we attribute deaminase efficiency increases primarily to improvements in base editor expression. Together, these findings establish that activity-independent

SE-PACE can improve the soluble expression of potentially growth-inhibiting proteins, in some cases without loss of activity.

Discussion

We have developed methods for the continuous evolution of proteins with improved soluble expression. Our approach quickly generates improved variants without the need for library cloning or discrete, time-intensive steps usually required to implement protein directed evolution methods. Through the use of a PACE-compatible split-intein pIII, which acts as an AND gate to bridge two orthogonal positive selections, SE-PACE is capable of evolving proteins in either the presence or absence of a simultaneous selection for protein activity.

The split-intein pIII AND gate allows two positive selections to take place in the same PACE experiment. In this work, we demonstrated its ability to bridge selections for target protein binding and soluble protein expression. In theory, however, this strategy can link any two orthogonal PACE selections. In addition, we show that the split-intein pIII can be used to constrain evolving SP populations to disfavor undesirable outcomes such as the formation of truncation products that cheat an activity-independent soluble expression selection.

The SE-PACE-evolved proteins did not always show improvements in T_m, and in some cases exhibited small decreases. As PACE is conducted at 37 °C and the starting proteins melted well above this temperature ($T_m > 55$ °C), no direct selection pressure was present for improved thermodynamic stability, and increases in T_m likely resulted from mutations that improved soluble expression. As thermodynamic stability and soluble expression are not always correlated, SE-PACE's selection for the latter property may not necessarily produce improvements in the former. With respect to expression yields, the evolved anti-GCN4 and anti-Htt scFvs reported here exhibited increases in the 2–6-fold range. In comparison, previous directed evolution efforts to improve expression of scFv and related antibody-derived proteins in E. coli have produced variants that increase yields by 2 to > 10-fold[38,46,47].

We have demonstrated that simultaneous selection for two different traits (here, soluble expression and binding activity) is possible in PACE through the evolution of scFvs with enhanced soluble expression in the E. coli cytosol. All dual-selection evolved scFvs tested showed binding to their target epitopes similar to that of the starting ScFv. In contrast, the activity-independent SE-PACE of rAPOBEC1 yielded some variants that lost enzymatic activity, as expected.

We have shown that single- and dual-selection SE-PACE can improve the expression of a diverse variety of proteins, including proteins of bacterial origin (MBP), engineered antibodies (scFvs), and eukaryotic enzymes (rAPOBEC1). It may also be interesting to apply SE-PACE to proteins that have emerged from long directed evolution campaigns, which often result in proteins with diminished stability[21,48]. By enforcing selection for protein activity using the split-intein pIII AND gate, it should be possible to preserve newly acquired protein function while improving soluble expression. Therefore, SE-PACE may also provide a way to rapidly reverse the deleterious effects of extensive mutation to improve the utility of previously engineered or evolved proteins.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41589-018-0121-5.

References

1. Braun, P. & LaBaer, J. High throughput protein production for functional proteomics. *Trends Biotechnol.* **21**, 383–388 (2003).
2. Socha, R. D. & Tokuriki, N. Modulating protein stability - directed evolution strategies for improved protein function. *FEBS J.* **280**, 5582–5595 (2013).
3. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* **5**, 172 (2014).
4. Terpe, K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **72**, 211–222 (2006).
5. Esposito, D. & Chatterjee, D. K. Enhancement of soluble protein expression through the use of fusion tags. *Curr. Opin. Biotechnol.* **17**, 353–358 (2006).
6. Giver, L., Gershenson, A., Freskgard, P.-O. & Arnold, F. H. Directed evolution of a thermostable esterase. *Proc. Natl Acad. Sci. USA* **95**, 12809–12813 (1998).
7. Cherry, J. R. et al. Directed evolution of a fungal peroxidase. *Nat. Biotechnol.* **17**, 379–384 (1999).
8. Sieber, V., Plückthun, A. & Schmid, F. X. Selecting proteins with improved stability by a phage-based method. *Nat. Biotechnol.* **16**, 955–960 (1998).
9. Khersonsky, O. et al. Directed evolution of serum paraoxonase PON3 by family shuffling and ancestor/consensus mutagenesis, and its biochemical characterization. *Biochemistry* **48**, 6644–6654 (2009).
10. Waldo, G. S., Standish, B. M., Berendzen, J. & Terwilliger, T. C. Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* **17**, 691–695 (1999).
11. Waldo, G. S. Genetic screens and directed evolution for protein solubility. *Curr. Opin. Chem. Biol.* **7**, 33–38 (2003).
12. Yuan, L., Kurek, I., English, J. & Keenan, R. Laboratory-directed protein evolution. *Microbiol. Mol. Biol. Rev.* **69**, 373–392 (2005).
13. Rakonjac, J. & Model, P. Roles of pIII in filamentous phage assembly. *J. Mol. Biol.* **282**, 25–41 (1998).
14. Badran, A. H. & Liu, D. R. Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nat. Commun.* **6**, 8425 (2015).
15. Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A. & Liu, D. R. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nat. Chem. Biol.* **10**, 216–222 (2014).
16. Dickinson, B. C., Leconte, A. M., Allen, B., Esvelt, K. M. & Liu, D. R. Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. *Proc. Natl Acad. Sci. USA* **110**, 9007–9012 (2013).
17. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011).
18. Dickinson, B. C., Packer, M. S., Badran, A. H. & Liu, D. R. A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. *Nat. Commun.* **5**, 5352 (2014).
19. Packer, M. S., Rees, H. A. & Liu, D. R. Phage-assisted continuous evolution of proteases with altered substrate specificity. *Nat. Commun.* **8**, 956 (2017).
20. Hubbard, B. P. et al. Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nat. Methods* **12**, 939–942 (2015).
21. Badran, A. H. et al. Continuous evolution of *Bacillus thuringiensis* toxins overcomes insect resistance. *Nature* **533**, 58–63 (2016).
22. Bryson, D. I. et al. Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat. Chem. Biol.* **13**, 1253–1260 (2017).
23. Grodberg, J. & Dunn, J. J. ompT encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. *J. Bacteriol.* **170**, 1245–1253 (1988).
24. Shis, D. L. & Bennett, M. R. Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. *Proc. Natl Acad. Sci. USA* **110**, 5028–5033 (2013).
25. Betton, J.-M. & Hofnung, M. Folding of a mutant maltose-binding protein of *Escherichia coli* which forms inclusion bodies. *J. Biol. Chem.* **271**, 8046–8052 (1996).
26. Chun, S.-Y., Strobel, S., Bassford, P. Jr & Randall, L. L. Folding of maltose-binding protein. Evidence for the identity of the rate-determining step in vivo and in vitro. *J. Biol. Chem.* **268**, 20855–20862 (1993).
27. Roodveldt, C., Aharoni, A. & Tawfik, D. S. Directed evolution of proteins for heterologous expression and stability. *Curr. Opin. Struct. Biol.* **15**, 50–56 (2005).
28. Schaerli, Y., Gili, M. & Isalan, M. A split intein T7 RNA polymerase for transcriptional AND-logic. *Nucleic Acids Res.* **42**, 12322–12328 (2014).
29. Rakonjac, J. Filamentous bacteriophages: biology and applications. in *Encyclopedia of Life Sciences* (John Wiley & Sons, Ltd., Chichester, UK (2012)).
30. Shah, N. H., Dann, G. P., Vila-Perelló, M., Liu, Z. & Muir, T. W. Ultrafast protein splicing is common among cyanobacterial split inteins: implications for protein engineering. *J. Am. Chem. Soc.* **134**, 11338–11341 (2012).
31. Zettler, J., Schütz, V. & Mootz, H. D. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS Lett.* **583**, 909–914 (2009).
32. Brissette, J. L., Weiner, L., Ripmaster, T. L. & Model, P. Characterization and sequence of the E*scherichia coli* stress-induced psp operon. *J. Mol. Biol.* **220**, 35–48 (1991).

33. Wojcik, J. et al. A potent and highly specific FN3 monobody inhibitor of the Abl SH2 domain. *Nat. Struct. Mol. Biol.* **17**, 519–527 (2010).
34. Holliger, P. & Hudson, P. J. Engineered antibody fragments and the rise of single domains. *Nat. Biotechnol.* **23**, 1126–1136 (2005).
35. Wörn, A. & Plückthun, A. Stability engineering of antibody single-chain Fv fragments. *J. Mol. Biol.* **305**, 989–1010 (2001).
36. Hanes, J., Jermutus, L., Weber-Bornhauser, S., Bosshard, H. R. & Plückthun, A. Ribosome display efficiently selects and evolves high-affinity antibodies in vitro from immune libraries. *Proc. Natl Acad. Sci. USA* **95**, 14130–14135 (1998).
37. der Maur, A. A. et al. Direct in vivo screening of intrabody libraries constructed on a highly stable single-chain framework. *J. Biol. Chem.* **277**, 45075–45085 (2002).
38. Waraho-Zhmayev, D., Meksiriporn, B., Portnoff, A. D. & DeLisa, M. P. Optimizing recombinant antibodies for intracellular function using hitchhiker-mediated survival selection. *Protein Eng. Des. Sel.* **27**, 351–358 (2014).
39. Lecerf, J.-M. et al. Human single-chain Fv intrabodies counteract in situ huntingtin aggregation in cellular models of Huntington's disease. *Proc. Natl Acad. Sci. USA* **98**, 4764–4769 (2001).
40. Gilbreth, R. N., Esaki, K., Koide, A., Sidhu, S. S. & Koide, S. A dominant conformational role for amino acid diversity in minimalist protein–protein interfaces. *J. Mol. Biol.* **381**, 407–418 (2008).
41. Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol. Cell* **10**, 1247–1253 (2002).
42. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
43. Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729 (2016).
44. Gaudelli, N. M. et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
45. Rees, H. A. et al. Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.* **8**, 15790 (2017).
46. Seo, M. J. et al. Engineering antibody fragments to fold in the absence of disulfide bonds. *Protein Sci.* **18**, 259–267 (2009).
47. Jespers, L., Schon, O., Famm, K. & Winter, G. Aggregation-resistant domain antibodies selected on phage by heat denaturation. *Nat. Biotechnol.* **22**, 1161–1165 (2004).
48. 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).

Author contributions
T.W. designed the research, performed the experiments, analyzed data, and wrote the manuscript. A.H.B. designed the research, analyzed data, and wrote the manuscript. T.P.H. performed and analyzed the mammalian cell base editing experiments. D.R.L. designed and supervised the research and wrote the manuscript.

Competing interests
The authors have filed a provisional patent application on the PACE system and related improvements.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41589-018-0121-5.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to D.R.L.

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

Methods

General methods. Antibiotics (Gold Biotechnology) were used at the following working concentrations: carbenicillin, 50 μg/mL; spectinomycin, 50 μg/mL; chloramphenicol, 25 μg/mL; kanamycin, 50 μg/mL; tetracycline, 10 μg/mL; streptomycin, 50 μg/mL. HyClone water (GE Healthcare Life Sciences) was used for PCR reactions and cloning. For all other experiments, water was purified using a MilliQ purification system (Millipore). Phusion U Hot Start DNA polymerase (Thermo Fisher Scientific) was used for all PCRs. Plasmids and SPs were cloned by USER assembly[21]. Genes were obtained as synthesized gBlock gene fragments from Integrated DNA Technologies or PCR amplified directly from *E. coli* genomic DNA. Plasmids were cloned and amplified using either Mach1 (Thermo Fisher Scientific) or Turbo (New England BioLabs) cells. Unless otherwise noted, plasmid or SP DNA was amplified using the Illustra TempliPhi 100 Amplification Kit (GE Healthcare Life Sciences) before Sanger sequencing. A full list of plasmids used in this work is given in Supplementary Table 5. A full list of reagents and equipment used in this work is given in Supplementary Table 6.

Preparation and transformation of chemically competent cells. Strain S2060 (ref. [20]) was used in all luciferase, phage propagation, and plaque assays, as well as in all PACE experiments. To prepare competent cells, an overnight culture was diluted 1,000-fold into 50 mL of 2xYT media (United States Biologicals) supplemented with tetracycline and streptomycin and grown at 37 °C with shaking at 230 r.p.m. to OD_{600} ~0.4–0.6. Cells were pelleted by centrifugation at 4,000g for 10 min at 4 °C. The cell pellet was then resuspended by gentle stirring in 2 mL of ice-cold LB media (United States Biologicals) and 2 mL of 2x TSS (LB media supplemented with 5% v/v DMSO, 10% w/v PEG 3350, and 20 mM $MgCl_2$) was added. The cell suspension was stirred to mix completely, aliquoted and frozen on dry ice, and stored at −80 °C until use.

To transform cells, 100 μL of competent cells thawed on ice was added to a prechilled mixture of plasmid (2 μL each; up to three plasmids per transformation) in 95 μL KCM solution (100 mM KCl, 30 mM $CaCl_2$, and 50 mM $MgCl_2$ in H_2O) and stirred gently with a pipette tip. The mixture was incubated on ice for 10 min and heat shocked at 42 °C for 75 s before 600 μL of SOC media (New England BioLabs) was added. Cells were allowed to recover at 37 °C with shaking at 230 r.p.m. for 1.5 h, streaked on 2xYT media + 1.5% agar (United States Biologicals) plates containing the appropriate antibiotics, and incubated at 37 °C for 16–18 h.

Luciferase assay. S2060 cells were transformed with the AP(s) and CP(s) of interest as described above. Overnight cultures of single colonies grown in 2xYT media supplemented with maintenance antibiotics were diluted 500-fold into DRM media[17] with maintenance antibiotics in a 96-well deep well plate (Axygen). The plate was sealed with a porous sealing film and grown at 37 °C with shaking at 230 r.p.m. for 4 h, whereupon the culture reached OD_{600} ~0.4–0.6. The cells were then induced with 100 ng/mL anhydrotetracycline (aTc; Fluka) and 5 μM arabinose (Gold Biotechnology) before incubation for an additional 1 h at 37 °C with shaking at 230 r.p.m. 120 μL of cells were transferred to a 96-well black-walled clear-bottomed plate (Costar), and then 600 nm absorbance and luminescence were read using an Infinite M1000 Pro microplate reader (Tecan). OD_{600}-normalized luminescence values were obtained by dividing raw luminescence by background-subtracted 600 nm absorbance. The background value was set to the 600 nm absorbance of wells containing DRM only.

Phage propagation assay. S2060 cells were transformed with the AP(s) of interest as described above. Overnight cultures of single colonies grown in 2xYT media supplemented with maintenance antibiotics were diluted 1,000-fold into DRM media with maintenance antibiotics and grown at 37 °C with shaking at 230 r.p.m. to OD_{600} ~0.4–0.6. Cells were then infected with SP at a starting titer of 5×10^4 pfu/mL. Cells were incubated for another 16–18 h at 37 °C with shaking at 230 r.p.m., then centrifuged at 8,000g for 2 min. The supernatant containing phage was removed and stored at 4 °C until use.

Plaque assay. S2060 cells were transformed with the AP(s) of interest as described above. Overnight cultures of single colonies grown in 2xYT media supplemented with maintenance antibiotics were diluted 1,000-fold into fresh 2xYT media with maintenance antibiotics and grown at 37 °C with shaking at 230 r.p.m. to OD_{600} ~0.6–0.8 before use. SP were serially diluted 100-fold (4 dilutions total) in H_2O. 150 μL of cells were added to 10 μL of each phage dilution, and to this 1 mL of liquid (55 °C) top agar (2xYT media + 0.6% agar) supplemented with 2% Bluo-gal (Gold Biotechnology) was added and mixed by pipetting up and down once. This mixture was then immediately pipetted onto one quadrant of a quartered Petri dish already containing 2 mL of solidified bottom agar (2xYT media + 1.5% agar, no antibiotics). After solidification of the top agar, plates were incubated at 37 °C for 16–18 h.

Phage-assisted continuous evolution. Unless otherwise noted, PACE apparatus, including host cell strains, lagoons, chemostats, and media, were all used as previously described[17,21]. To reduce the likelihood of contamination with gIII-encoding recombinant SP, phage stocks were purified as previously described[21].

Chemically competent S2060s were transformed with AP(s) and MP6 or DP6 as described above, plated on 2xYT media + 1.5% agar supplemented with 25 mM glucose (to prevent induction of mutagenesis) in addition to maintenance antibiotics, and grown at 37 °C for 18–20 h. Four colonies were picked into 1 mL DRM each in a 96-well deep-well plate, and this was diluted five-fold eight times serially into DRM. The plate was sealed with a porous sealing film and grown at 37 °C with shaking at 230 r.p.m. for 16–18 h. Dilutions with OD_{600} ~0.4–0.8 were then used to inoculate a chemostat containing 80 mL DRM. The chemostat was grown to OD_{600} ~0.8–1.0, then continuously diluted with fresh DRM at a rate of ~1.5 chemostat volumes/h as previously described[17]. The chemostat was maintained at a volume of 60–80 mL.

Prior to SP infection, lagoons were continuously diluted with culture from the chemostat at 1 lagoon vol/h and pre-induced with 10 mM arabinose for at least 2 h. If DP6 was used, the lagoons were also pre-induced with aTc. Lagoons were infected with SP at a starting titer of 10^6 pfu/mL and maintained at a volume of 15 mL. Samples (500 μL) of the SP population were taken at indicated times from lagoon waste lines. These were centrifuged at 8,000g for 2 min, and the supernatant was passed through a 0.22 μm PVDF Ultrafree centrifugal filter (Millipore) and stored at 4 °C. Lagoon titers were determined by plaque assays using S2060 cells transformed with pJC175e[15]. For Sanger sequencing of lagoons, single plaques were PCR amplified using primers AB1793 (5′-TAATGGAAACTTCCTCATGAAAAAGTCTTTAG) and AB1396 (5′-ACAGAGAGAATAACATAAAAACAGGGAAGC), both of which anneal to regions of the phage backbone flanking the evolving gene of interest. Generally, eight plaques were picked and sequenced per lagoon.

Continuous flow competition of MBP and G32D+I33P. Host cells transformed with pTW006aP1a were maintained in a 40 mL chemostat. Lagoons were infected with a 1,000:1 ratio of SP02a (G32D + I33P) to SP01a (MBP). No inducers were used in this continuous flow experiment. The SP pool was monitored by PCR using primers AB1793 (5′-TAATGGAAACTTCCTCATGAAAAAGTCTTTAG), which anneals to the phage backbone 5′ of the MBP–T7n insert, and TW0023 (5′-ACCACCAGATCCACCCUTGGTGATACGAGTCTGCG), a USER primer that anneals to the 3′ end of MBP. The purified PCR product (200 ng) was digested with BsaWI (New England BioLabs), which cuts a ACC|GGA motif in MBP that is absent in the G32D + I33P mutant, according to manufacturer's protocol.

Evolution of MBP Y283D and G32D+I33P. Host cells transformed with pTW006aP1a and MP6 were maintained in a 40 mL chemostat. Lagoons were infected with SP02a (G32D + I33P) or SP02b (Y293D). Upon infection, lagoon dilution rates were decreased to 0.5 volume/h. For G32D + I33P, the lagoon dilution rate was maintained at 0.5 vol/h throughout the experiment. For Y283D, the lagoon dilution rate was increased to 1 vol/h at 22 h. Both experiments ended at 72 h.

Evolution of HA4 Y87A monobody. Host cells transformed with pTW048a3, pTW051b2, and DP6 were maintained in a 40 mL chemostat. The lagoon was cycled at 1 vol/h with 10 mM arabinose and 20 ng/mL aTc for 3 h before infection with SP10b2. Upon infection, lagoon dilution rates were maintained at 1 vol/h. The lagoon dilution rate was increased to 1.5 vol/h at 93 h and 2 vol/h at 119 h. The experiment ended at 122 h.

Evolution of anti-GCN4 scFv Ωg. Host cells transformed with pTW055a3, pTW051b4, and MP6 were maintained in a 40 mL chemostat. Lagoons were infected with SP16c3. Upon infection, lagoon dilution rates were decreased to 0.5 volume/h. After 19 h, the lagoon dilution rate was increased to 1 volume/h. The experiment ended at 72 h.

Evolution of disulfide-free anti-GCN4 scFvs. Host cells transformed with pTW055a3, pTW051b4, and MP6 were maintained in a 40 mL chemostat. Lagoons were infected with SP39b or SP39d. Upon infection, lagoon dilution rates were decreased to 0.5 vol/h. For SP39b, the lagoon dilution rate was increased to 1 vol/h after 19 h, 1.5 vol/h after 44 h, 2 vol/h after 68 h, and 3 vol/h after 90 h. For SP39d, the lagoon dilution rate was increased to 1 vol/h after 44 h, 1.5 vol/h after 68 h, and 2 vol/h after 90 h. The experiment ended at 120 h.

Evolution of anti-htt scFv C4. Host cells transformed with pTW074c, pTW051b4, and MP6 were maintained in an 80 mL chemostat. Lagoons were infected with SP24a3. Upon infection, lagoon dilution rates were decreased to 0.5 vol/h. The lagoon dilution rate was increased to 1 vol/h at 42 h, 1.5 vol/h at 70 h, 2 vol/h at 90 h, and 3 vol/h at 114 h. The experiment ended at 120 h.

Further evolution of anti-htt scFv C4. Host cells transformed with pTW074c, pTW051d2, and MP6 were maintained in a 40 mL chemostat. Lagoons were infected with the 120 h SP pool from the previous PACE. Upon infection, lagoon dilution rates were decreased to 0.5 vol/h. The lagoon dilution rate was increased to 1 vol/h at 18 h, 1.5 vol/h at 42 h, 2 vol/h at 69 h, and 3 vol/h at 90 h. The experiment ended at 114 h.

Evolution of MBP G32D+I33P with affinity tag strategy. Host cells transformed with pTW084b, pTW051b4, and MP6 were maintained in an 80 mL chemostat. Lagoons were infected with SP27b. Upon infection, lagoon dilution rates were decreased to 0.5 vol/h. For lagoon 1, the dilution rate was increased to 1 vol/h at 44 h and 2 vol/h at 69 h. For lagoon 2, the dilution rate was increased to 1 vol/h at 44 h, 1.5 vol/h at 69 h, and 2 vol/h at 88 h. The experiment ended at 112 h.

Evolution of rAPOBEC1 with affinity tag strategy. Host cells transformed with pTW084b, pTW051d, and DP6 were maintained in a 40 mL chemostat. The lagoon was cycled at 1 vol/h with 10 mM arabinose and 20 ng/mL aTc for 4 h before infection with SP30. Upon infection, lagoon dilution rates were decreased to 0.5 vol/h. At 18 h, the aTc concentration was decreased to 0 ng/mL. The lagoon dilution rate was increased to 1 vol/h at 28 h and 1.5 vol/h at 66 h. The experiment ended at 74 h.

Evolution was continued on host cells transformed with pTW084b, pTW051d, and MP6 in an 80 mL chemostat. Lagoons were infected with a 1:1 ratio of 35.1 and 35.2, SP clones isolated from the previous experiment. Upon infection, lagoon dilution rates were maintained at 1 vol/h. The lagoon dilution rate was increased to 1.5 vol/h at 18 h, 2 vol/h at 66 h, and 3 vol/h at 90 h. The experiment ended at 112 h.

Evolution was continued on host cells transformed with pTW084b, pTW051b4, and MP6 in an 80 mL chemostat. Lagoons were infected with either 36.1 or 36.2, SP clones isolated from the previous experiment. Upon infection, lagoon dilution rates were decreased to 0.5 vol/h. The lagoon dilution rate was increased to 1 vol/h at 17 h, 1.5 vol/h at 41 h, and 2 vol/h at 66 h. At 96 h, the chemostat was replaced with a fresh 80 mL chemostat containing host cells transformed with pTW084b, pTW051d2, and MP6, and the lagoon dilution rate was decreased to 1 vol/h. The lagoon dilution rate was increased to 1.5 vol/h at 115 h and 2 vol/h at 168 h. The experiment ended at 184 h.

Small-scale protein expression. BL21 DE3 cells (New England BioLabs) were transformed with the expression plasmids (EPs) of interest according to manufacturer protocol. Overnight cultures of single colonies grown in 2xYT media supplemented with maintenance antibiotics were diluted 1,000-fold into fresh 2xYT media (2 mL) with maintenance antibiotics and grown at 37 °C with shaking at 230 r.p.m. to OD_{600} ~0.4–0.6 before induction with 1 mM isopropyl-β-D-thiogalactoside (IPTG; Gold Biotechnology) or rhamnose (Gold Biotechnology). Cells were grown for a further 3 h at 37 °C with shaking at 230 r.p.m. Cells from 1.4 mL of culture were isolated by centrifugation at 8,000g for 2 min. The resulting pellet was resuspended in 150 μL B-per reagent (Thermo Fisher Scientific) supplemented with protease inhibitor cocktail (Roche) and incubated on ice for 10 min before centrifugation at 16,000g for 2 min. The supernatant was collected as the soluble fraction, and the resulting pellet was resuspended in an additional 150 μL B-per reagent to obtain the insoluble fraction. To 37.5 μL of each fraction was added 12.5 μL 4×Laemmli sample loading buffer (Bio-Rad) containing 2 mM dithiothreitol (DTT; Sigma-Aldrich). After vortexing, the fractions were incubated at 95 °C for 10 min. 12 μL of each soluble fraction and 6 μL of each insoluble fraction was loaded into the wells of a Bolt 4–12% Bis–Tris Plus (Thermo Fisher Scientific) pre-cast gel. 6 μL of Precision Plus Protein Dual Color Standard (Bio-Rad) was used as a reference. Samples were separated by electrophoresis at 180 V for 35 min in Bolt MES SDS running buffer (Thermo Fisher Scientific). Gels were stained with InstantBlue reagent (Expedeon) for 1 h to overnight, then washed several times with H_2O before imaging with a G:Box Chemi XRQ (Syngene). Band densities were quantified using ImageJ and normalized to reference bands to control for protein loading.

Western blot analysis. SDS–PAGE was performed as described above. Transfer to a PVDF membrane was performed using an iBlot 2 Gel Transfer Device (Thermo Fisher Scientific) according to the manufacturer's protocols. The membrane was blocked in Odyssey Blocking Buffer (LI-COR) in PBS for 1 h at room temperature, then incubated with mouse anti-6×His (Abcam ab18184; 1:2,000 dilution) and rabbit anti-GroEL (Sigma-Aldrich G6532; 1:20,000 dilution) in SuperBlock Blocking Buffer (Thermo Fisher Scientific) overnight at 4 °C. The membrane was washed 3× with TBST (TBS + 0.5% Tween-20) for 10 min each at room temperature, then incubated with IRDye-labeled secondary antibodies goat anti-mouse 680RD (LI-COR 926-68070) and donkey anti-rabbit 800CW (LI-COR 926-32213) diluted 1:5,000 in SuperBlock for 1 h at room temperature. The membrane was washed as before and then imaged using an Odyssey Imaging System (LI-COR).

Medium-scale protein expression and purification. BL21 DE3 cells were transformed with the EPs of interest according to manufacturer protocol. Overnight cultures of single colonies grown in LB or 2xYT media supplemented with maintenance antibiotics were diluted 1,000-fold into fresh LB or 2xYT media (250 mL) with maintenance antibiotics and grown at 37 °C with shaking at 230 r.p.m. to OD_{600} ~0.4–0.6. Cells were chilled on ice for 1 h, then induced with 1 mM IPTG and grown for a further 16–18 h at 16 °C with shaking at 200 r.p.m. For 37 °C post-induction growth, the cold shock step was omitted and the cells were grown for a further 3 h at 37 °C with shaking at 200 r.p.m. after induction with 1 mM IPTG. Cells were isolated by centrifugation at 8,000g for 10 min.

The resulting pellet was resuspended in 4 mL B-per reagent supplemented with EDTA-free protease inhibitor cocktail (Roche) and incubated on ice for 20 min before centrifugation at 12,000g for 15 min. The supernatant was decanted into a 15 mL conical tube and incubated with 250 μL of TALON Cobalt (Clontech) resin at 4 °C with constant agitation for 1–2 h, after which the resin was isolated by centrifugation at 500g for 5 min. The supernatant was decanted, and the resin resuspended in 2 mL binding buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 7.8) and transferred to a column. The resin was washed 4× with 1 mL binding buffer before protein was eluted with 1 mL of binding buffer containing increasing concentrations of imidazole (50–250 mM in 50 mM increments). The fractions were analyzed by SDS–PAGE for purity. Combined fractions were buffer-exchanged with TBS (20 mM Tris–Cl, 500 mM NaCl, pH 7.5) and concentrated using an Amicon Ultra-15 centrifugal filter unit (10,000 molecular weight cutoff; Millipore), then stored at 4 °C until further use. Proteins were quantified using Quick Start Bradford reagent (Bio-Rad) using BSA standards (Bio-Rad).

Medium-scale BE3 expression and purification. BE3 variants were expressed and purified as described previously[45] with a few modifications. BL21 DE3 cells were transformed with the BE3-expressing EPs of interest according to manufacturer protocol. Overnight cultures of single colonies grown in 2xYT media supplemented with maintenance antibiotics were diluted 1,000-fold into fresh 2xYT media (200 mL) with maintenance antibiotics and grown at 37 °C with shaking at 230 r.p.m. to OD_{600} ~0.7–0.8. Cells were chilled on ice for 1 h, then induced with IPTG and grown for a further 16–18 h at 16 °C with shaking at 200 r.p.m. Cells were isolated by centrifugation at 8,000g for 15 min. The resulting pellet was resuspended in 8 mL high salt buffer (100 mM Tris–Cl, 1 M NaCl, 5 mM tris(2-carboxyethyl)phosphine (TCEP; Gold Biotechnology), 20% glycerol, pH 8.0) supplemented with EDTA-free protease inhibitor cocktail and 1 mM phenylmethane sulfonyl fluoride (PMSF; Sigma-Aldrich). Cells were sonicated on ice (3 s on/3 s off; 6 min total) and the lysate centrifuged at 16,000g for 15 min. The supernatant was decanted into a 15 mL conical tube and incubated with 500 μL of TALON Cobalt resin at 4 °C with constant agitation for 1 h, after which the resin was isolated by centrifugation at 500g for 5 min. The resin was washed 5× with 1 mL high salt buffer, then eluted with 1 mL of elution buffer (100 mM Tris–Cl, 500 mM NaCl, 5 mM TCEP, 200 mM imidazole, 20% glycerol, pH 8.0). The isolated protein was then buffer-exchanged with medium salt buffer (100 mM Tris–Cl, 500 mM NaCl, 5 mM TCEP, 20% glycerol, pH 8.0) and concentrated using an Amicon Ultra-15 centrifugal filter unit (100,000 molecular weight cutoff). Proteins were quantified using Quick Start Bradford reagent using BSA standards.

Protein melt temperature assay. Protein melt temperatures were determined using a Protein Thermal Shift Dye Kit (Life Technologies) according to manufacturer's protocols. Fluorescence was monitoring using a CFX96 Real-Time PCR Detection System (Bio-Rad).

ELISA. *Anti-GCN4 ELISA.* A MaxiSorp 96-well plate (Nunc) was coated with a 10 μg/mL solution of purified antibody fragments in TBS (20 mM Tris–Cl, 500 mM NaCl, pH 7.5), 50 μL per well, and incubated overnight at 4 °C. The wells were washed 3× with TBST (TBS + 0.5% Tween-20) and blocked with 1% w/v bovine serum albumin (BSA; Sigma-Aldrich) in TBS for 1–2 h. This and all subsequent incubations were performed at room temperature. The blocking solution was removed and the wells were incubated with serial dilutions of purified MBP-TEV-GCN4 in TBS for 1 h. The wells were washed 3× with TBST, then incubated with a 1:5,000 dilution of HRP-conjugated anti-MBP (Abcam ab49923) in TBS for 1 h. The wells were washed 3× with TBST, then developed with 1-Step Ultra TMB-ELISA Substrate (Thermo Fisher Scientific) for 5 min. The reaction was quenched with addition of 2 M H_2SO_4 before 450 nm absorbance was read using a Tecan Infinite M1000 Pro microplate reader.

Anti-htt ELISA. A MaxiSorp 96-well plate was coated with a 50 μg/mL solution of purified antibody fragments in TBS, 50 μL per well, and incubated overnight at 4 °C. The wells were washed and blocked as above. The blocking solution was removed, and the wells were incubated with serial dilutions of biotinylated htt peptide (MATLEKLMKAFESLKSFK(Biotin)-NH₂; New England Peptide) in TBS for 1 h. The wells were washed 3× with TBST, then incubated with a 1:10,000 dilution of HRP-conjugated streptavidin (BioLegend) in TBS for 1 h. The wells were washed 3× with TBST, and then developed and quenched, and the absorbance read as above.

Rifampicin resistance assay. BL21 DE3 cells were transformed with the EPs of interest according to manufacturer protocol. Overnight cultures of single colonies grown in DRM media supplemented with maintenance antibiotics were diluted 1,000-fold into DRM media with maintenance antibiotics in a 96-well deep well plate. The plate was sealed with a porous sealing film and grown at 37 °C with shaking at 230 r.p.m. until the culture reached OD_{600} ~0.4. The cells were then either induced with 5 mM rhamnose or repressed with 5 mM glucose before incubation for an additional 16–18 h at 37 °C with shaking at 230 r.p.m. Cultures were serially diluted on 2xYT + 1.5% agar plates supplemented with

50 μg/mL spectinomycin, 100 μg/mL rifampin (Alfa Aesar), and 25 mM glucose. The total number of colony-forming units (cfus) was determined by serially diluting the same cultures on 2xYT + 1.5% agar plates supplemented with 50 μg/mL spectinomycin and 25 mM glucose. Plates were grown at 37 °C for 16–18 h. The surviving colonies on the plates containing rifampin were counted and this number was normalized to the total cfu count.

Chloramphenicol resistance assay. Chemically competent S2060s were transformed with selection plasmid[44] and the base editor EPs of interest as described above and streaked on 1.5% agar plates supplemented with 50 μg/mL carbenicillin, 50 μg/mL spectinomycin, and 25 mM glucose. Overnight cultures of single colonies grown in DRM media supplemented with maintenance antibiotics were diluted 500-fold into DRM media with maintenance antibiotics and grown at 37 °C with shaking at 230 r.p.m. for until the culture reached OD_{600} ~0.5. The cells were then induced with 10 mM arabinose and incubated for an additional 16–18 h at 37 °C with shaking at 230 r.p.m. Cultures were serially diluted on 2xYT + 1.5% agar plates supplemented with 32 or 64 μg/mL chloramphenicol and 25 mM glucose (to repress further expression). The total number of cfus was determined by serially diluting the same cultures on 2xYT + 1.5% agar plates supplemented with 50 μg/mL carbenicillin, 50 μg/mL spectinomycin, and 25 mM glucose. Plates were grown at 37 °C for 16–18 h. The surviving colonies on the plates containing chloramphenicol were counted, and this number was normalized to the total cfu count.

HEK293T transfection and genomic DNA extraction. HEK293T cells (ATCC CRL-3216) maintained in Dulbecco's Modified Eagle's Medium plus GlutaMax (Thermo Fisher Scientific) supplemented with 10% (v/v) FBS at 37 °C with 5% CO_2 were seeded on 48-well poly-D-lysine coated plates (Corning). 16–20 h after seeding, cells were transfected at approximately 80–85% confluency as previously described[44]. Cells were cultured for 3 d post-transfection before genomic DNA was extracted as previously described[44].

High-throughput sequencing of genomic DNA. Genomic sites were amplified with primers targeting the region of interest and the appropriate universal Illumina forward and reverse adapters. Primer pairs for the first round of PCR (PCR 1) were the same as reported previously[42]. 25 μL scale PCR 1 reactions used 1.25 μL each of 10 μM forward and reverse primers and 0.5 μL genomic DNA extract, all diluted to 12.5 μL with nuclease-free water, and 12.5 μL Phusion U Green Multiplex PCR MasterMix (Thermo Fisher Scientific). PCR 1 conditions: 98 °C for 2 min, then 30 cycles of (98 °C for 15 s, 61 °C for 20 s, 72 °C for 15 s), followed by a final 72 °C extension for 2 min. PCR products were verified by comparison with DNA standards (Quick-Load 2-Log Ladder; New England BioLabs) on a 2% agarose gel supplemented with ethidium bromide. Unique Illumina barcoding primers that

anneal to the universal Illumina adaptor region were subsequently appended to each PCR 1 sample in a second PCR reaction (PCR 2). PCR 2 reactions used 1.25 μL each of 10 μM forward and reverse Illumina barcoding primers and 1 μL of unpurified PCR 1 reaction product, all diluted to 12.5 μL with nuclease-free water, and 12.5 μL Phusion U Green Multiplex PCR MasterMix (Thermo Fisher Scientific). PCR 2 conditions: 98 °C for 2 min, then 12 cycles of (98 °C for 15 s, 61 °C for 20 s, 72 °C for 20 s), followed by a final 72 °C extension for 2 min. PCR products were pooled and purified by electrophoresis with a 2% agarose gel using a Monarch DNA Gel Extraction Kit (New England BioLabs) eluting with 30 μL H_2O. DNA concentration was quantified with the KAPA Library Quantification Kit-Illumina (KAPA Biosystems) and sequenced on an Illumina MiSeq instrument (paired-end read – R1: 220 cycles, R2: 0 cycles) according to the manufacturer's protocols.

General HTS data analysis. Sequencing reads were demultiplexed in MiSeq Reporter (Illumina). Alignment of amplicon sequences to a reference sequence was performed as previously described using a custom MATLAB script[42]. In brief, the Smith–Waterman algorithm was used to align sequences without indels to a reference sequence; bases with a quality score of less than 30 were converted to 'N' to prevent base miscalling as a result of sequencing error. Indels were quantified separately using a modified version of a previously described MATLAB script in which sequencing reads with more than half the base calls below a quality score of Q30 were filtered out[44]. Indels were counted as reads which contained insertions or deletions of greater than or equal to 1 bp within a 30 bp window surrounding the predicted Cas9 cleavage site.

Base editing values are representative of $N = 4$ independent biological replicates collected over different days, with the mean ± s.e.m. shown. Base editing values are reported as a percentage of the number of reads with cytidine mutagenesis over the combined number of aligned reads and indel-containing reads.

Statistics. Information regarding the number and type of replicates, along with the type of error depicted, is provided in the figure legends.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Code availability. The custom MATLAB script used in this work can be found in the supplementary information of ref. [39].

Data availability. High-throughput sequencing data for Fig. 5f and Supplementary Fig. 19 have been deposited in the NCBI Sequence Read Archive database under accession code SRP152041. Selection plasmids used in this study will be available through Addgene. Other data and materials are available upon reasonable request.

natureresearch

Corresponding author(s): David R. Liu

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see <u>Authors & Referees</u> and the <u>Editorial Policy Checklist</u>.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The <u>exact sample size</u> (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including <u>central tendency</u> (e.g. means) or other basic estimates (e.g. regression coefficient) AND <u>variation</u> (e.g. standard deviation) or associated <u>estimates of uncertainty</u> (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on <u>statistics for biologists</u> may be useful.

Software and code

Policy information about <u>availability of computer code</u>

Data collection	None used
Data analysis	ImageJ v 1.50i, Matlab R2017a (9.2.0) with previously published scripts

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research <u>guidelines for submitting code & software</u> for further information.

Data

Policy information about <u>availability of data</u>

All manuscripts must include a <u>data availability statement</u>. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

High-throughput sequencing data for Figures 5f and SI Figure 19 have been deposited in the NCBI Sequence Read Archive database under accession code SRP152041. Selection plasmids used in this study will be available through Addgene. Other materials are available upon reasonable request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences

Study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were determined based on literature precedence for directed evolution.
Data exclusions	No data excluded
Replication	With the exception of evolutions (where replication is not standard in the field), some protein expressions (noted in the figure legends -- SI Figures 10c, 15a,b), and the plaque assay in main text Figure 3c, all experiments were replicated at least once. All attempts at replication were successful.
Randomization	Bacteria and/or mammalian cells used for these experiments were grown under identical conditions, so randomization was not used.
Blinding	Bacteria and/or mammalian cells used for these experiments were grown under identical conditions, so blinding was not used.

Materials & experimental systems

Policy information about availability of materials

n/a	Involved in the study
☐	☒ Unique materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Research animals
☒	☐ Human research participants

Unique materials

Obtaining unique materials	Selection plasmids used in this study will be available through Addgene. Other materials are available upon reasonable request.

Antibodies

Antibodies used	Antibodies were used for western blotting analysis. The following antibodies were used: Anti-6xHis, mouse monoclonal, abcam ab18184 (http://www.abcam.com/6x-his-tag-antibody-hish8-ab18184.html) lot# GR3185677-4, clone HIS.H8, 1:2000 dilution Anti-GroEL, rabbit polyclonal, Sigma-Aldrich G6532 (https://www.sigmaaldrich.com/catalog/product/sigma/g6532?lang=en&region=US) lot# 116M4854V, 1:20,000 dilution Anti-MBP HRP conjugated, mouse monoclonal, abcam ab49923 (http://www.abcam.com/maltose-binding-protein-antibody-mbp-17-hrp-ab49923.html) lot# GR251034-5, clone MBP-17, 1:5000 dilution Anti-mouse 680RD secondary, goat polyclonal, LI-COR 926-68070 lot# C70427-05, 1:5000 diluton Anti-rabbit 800CW secondary, donkey polyclonal, LI-COR 926-32213 lot# C50512-05, 1:5000 dilution
Validation	Anti-6xHis, mouse monoclonal, abcam ab18184: validated for western blotting by manufacturer (blotting against rRab27a-His in whole cell lysate) Anti-GroEL, rabbit polyclonal, Sigma-Aldrich G6532: validated for western blotting by manufacturer (blotting against GroEL in E. coli bacterial extract) Anti-MBP HRP conjugated, mouse monoclonal, abcam ab49923: validated for ELISA by manufacturer (ELISA against MBP) Anti-mouse 680RD secondary, goat polyclonal, LI-COR 926-68070: validated for western blotting by manufacturer (dot blot against serum proteins) Anti-rabbit 800CW secondary, donkey polyclonal, LI-COR 926-32213 lot# C50512-05, 1:5000 dilution; validated for western blotting by manufacturer (dot blot against serum proteins)

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T (ATCC)
Authentication	Cells were authenticated by the supplier using STR analysis.
Mycoplasma contamination	HEK293T cells tested negative for mycoplasma.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Method-specific reporting

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ Magnetic resonance imaging