(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 May 2016 (19.05.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/077052 A2

(51) International Patent Classification: *C12N 15/70* (2006.01)

(21) International Application Number: PCT/US2015/057012

(22) International Filing Date: 22 October 2015 (22.10.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/067,194 22 October 2014 (22.10.2014) US

(71) Applicant: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

(72) Inventors: **LIU, David, R.**; 3 Whitman Circle, Lexington, MA 02420 (US). **DICKINSON, Bryan**; 5500 S. Harper Avenue, Chicago, IL 60637 (US). **PACKER, Michael, S.**; 202 Franklin Street, Apt. 3L, Cambridge, MA 02139 (US). **BADRAN, Ahmed, Hussein**; 84 Line Street, Apt. 3, Somerville, MA 02143 (US).

(74) Agent: **BAKER, C., Hunter**; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

(54) Title: EVOLUTION OF PROTEASES

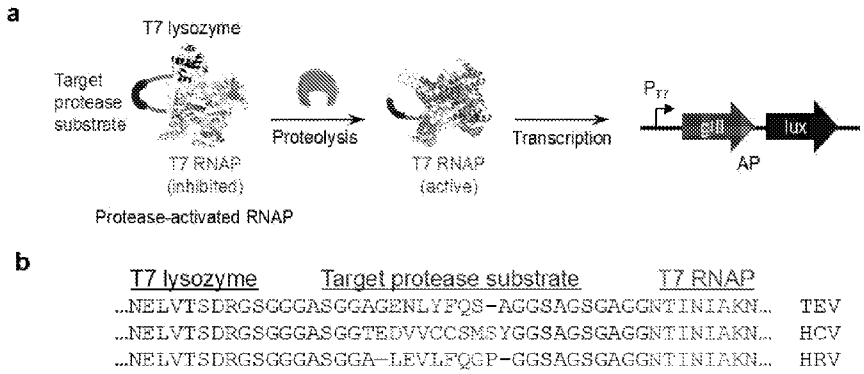


FIG. 1

(57) Abstract: Some aspects of this disclosure provide methods for phage-assisted continuous evolution (PACE) of proteases. Some aspects of this invention provide methods for evaluating and selecting protease inhibitors based on the likelihood of the emergence of resistant proteases as determined by the protease PACE methods provided herein. Some aspects of this disclosure provide strategies, methods, and reagents for protease PACE, including fusion proteins for translating a desired protease activity into a selective advantage for phage particles encoding a protease exhibiting such an activity and improved mutagenesis-promoting expression constructs. Evolved proteases that recognize target cleavage sites which differ from their canonical cleavage site are also provided herein.

WO 2016/077052 A2

EVOLUTION OF PROTEASES

RELATED APPLICATION

[0001] This application claims priority under 35 U.S.C. § 119(e) to U.S. provisional patent application, U.S.S.N. 62/067,194, filed October 22, 2014, which is incorporated herein by reference.

GOVERNMENT SUPPORT

[0002] This invention was made with Government support under grant T32 GM008313 awarded by the National Institutes of Health (NIH) and grants HR0011-11-2-0003 and N66001-12-C-4207 awarded by the Defense Advanced Research Projects Agency (DARPA). The Government has certain rights in this invention.

BACKGROUND

[0003] Proteases are ubiquitous enzymes that play important roles in many aspects of cell and tissue biology. Proteases can also be harnessed for biotechnological and biomedical applications. Among the more than 600 naturally occurring proteases that have been described[1] are enzymes that have proven to be important catalysts of industrial processes, essential tools for proteome analysis, and life-saving pharmaceuticals[2-5]. Recombinant human proteases including thrombin, factor VIIa, and tissue plasminogen activator are widely used drugs for the treatment of blood clotting diseases[4]. In addition, the potential of protease-based therapeutics to address disease in a manner analogous to that of antibody drugs,[6,7] but with catalytic turnover, has been recognized for several decades[4,8].

[0004] Natural proteases, however, typically target only a narrowly defined set of substrates, limiting their therapeutic potential. The directed evolution of proteases in principle could generate enzymes with tailor-made specificities, but laboratory-evolved proteases are frequently non-specific, weakly active, or only modestly altered in their substrate specificity, limiting their utility[9-14].

[0005] In addition to their importance as current and future therapeutic agents, proteases have also proven to be major drug targets for diseases including cardiovascular illness, infectious disease, and cancer[15,16]. While drug specificity and potency are characterized and optimized during pre-clinical studies, the evolution of drug resistance is often not well understood until it arises in patients, despite the strong relationship between drug resistance vulnerability and a lack of therapeutic efficacy. For example, resistance to

HIV and HCV protease inhibitors can arise in as few as two days of clinical use[17] and frequently leads to viral rebound and poor treatment outcomes[18-21]. The speed with which drug resistance can arise in the clinic endangers patients and puts years of drug development efforts prior to such a determination at risk.

[0006] Characterizing the potential of protease inhibitors to be overcome by the evolution of drug resistance using methods such as mammalian cell culture, animal models, or yeast display-based laboratory evolution is time- and labor-intensive[22,23]. As a result, identifying drug resistance vulnerabilities of early-stage preclinical candidates is not common practice.

SUMMARY

[0007] The laboratory evolution of protease enzymes has the potential to generate proteases with therapeutically relevant specificities, and to assess the vulnerability of protease inhibitor drug candidates to the evolution of drug resistance. Some aspects of this disclosure describe a system for the continuous directed evolution of proteases using phage-assisted continuous evolution (PACE) that links the proteolysis of a target peptide to phage propagation through a protease-activated RNA polymerase (PA-RNAP). Some aspects of this disclosure describe the engineering of an expression system that directly links protease activity to phage propagation, for example, by fusing a transcriptional activator to an inhibitory domain via a protease-cleavable linker. The continuous evolution technology provided herein is useful for evolving proteases with altered substrate preferences, specificities, and cleavage efficiencies. In addition, the technology provided herein can also be used to analyze how therapeutically targeted proteases acquire resistance to therapeutic protease inhibitors, to evaluate a candidate protease inhibitor regarding the potential of the target protease to develop resistance, and to design improved protease inhibitors that are not rendered ineffective simply as a result of one or two mutations in their target protease.

[0008] Phage-assisted continuous evolution (PACE) can serve as a rapid, high-throughput method to evolve a protease to reveal resistance to protease inhibitor drug candidates, analogous to previous uses of stepwise protein evolution to study antibiotic resistance[24]. One advantage of the PACE technology described herein is that both the time and human effort required to evolve a protease or to evaluate the likelihood of protease inhibitor resistance to develop are dramatically decreased as compared to conventional iterative evolution methods.

[0009] The general concept of PACE technology has been described, for example in International PCT Application, PCT/US2009/056194, filed September 8, 2009, published as WO 2010/028347 on March 11, 2010; International PCT Application, PCT/US2011/066747, filed December 22, 2011, published as WO 2012/088381 on June 28, 2012; and U.S. Application, U.S.S.N. 13/922,812, filed June 20, 2013, each of which is incorporated herein by reference. During PACE, a phage vector carrying a gene encoding the protease of interest replicates in a flow of host cells through a fixed-volume vessel (a "lagoon"). For example, in some embodiments of PACE described herein, a population of bacteriophage vectors replicates in a continuous flow of bacterial host cells through the lagoon, wherein the flow rate of the host cells is adjusted so that the average time a host cell remains in the lagoon is shorter than the average time required for host cell division, but longer than the average life cycle of the vector, *e.g.*, shorter than the average M13 bacteriophage life cycle. As a result, the population of vectors replicating in the lagoon can be varied by inducing mutations, and then enriching the population for desired variants by applying selective pressure, while the host cells do not effectively replicate in the lagoon.

[0010] Once the appropriate selective pressure is applied, for example, by linking a desired protease activity to the production of a gene for the generation of infectious phage particles as described in more detail elsewhere herein, phage with genes encoding proteases with the desired target activity acquire a selective advantage and thus replicate preferentially. For example, in the context of M13 phage PACE, selective pressure can be engineered by linking a desired target activity to the production of pIII, an essential component in the bacteriophage life cycle[25]. Because the lagoon is continuously diluted by a constant influx of fresh host cells, phage encoding inactive variants that do not trigger expression of the gene required for the generation of infectious phage particles are not packaged into infectious phage particles and are rapidly diluted out of the lagoon. Dilution occurs faster than cell division but slower than phage replication, ensuring that mutations only accumulate in the phage genome. Because evolution during PACE can take place continuously without researcher intervention, hundreds of rounds of evolution can be performed in a single PACE experiment within days or weeks, as compared to the performance of typically a single round of iterative evolution methods per day or per week.

[0011] Some aspects of this disclosure are based on the recognition that PACE is well-suited for the directed evolution of proteases, which, depending on the circumstances, may require many successive mutations to remodel complex networks of contacts with polypeptide substrates[26,27]. Moreover, the speed and efficiency of PACE may enable the rapid

identification of mutations that confer resistance to protease inhibitors. Previously, PACE has been reported to evolve RNA polymerase enzymes[25,28-30]. The development of new reagents and vectors that efficiently link protease activity to phage replication efficiency as described herein enable the rapid evolution of proteases with desired characteristics and the identification of protease mutations that confer resistance to protease inhibitors. In addition, this disclosure provides method for evaluating candidate protease inhibitors for the likelihood of a target protease developing a resistance, which may hamper clinical application of the inhibitor. Such evaluations cannot readily be performed with conventional techniques. As a result, resistances of proteases that are targeted by therapeutic protease inhibitors are often only recognized after an inhibitor is used in the clinic. The technology provided herein can be used to interrogate resistance formation to protease inhibitors before a candidate inhibitor is developed in a clinical setting, thus streamlining the evaluation of candidate protease inhibitors for drug development.

[0012] The utility of the PACE technology provided herein in the identification of resistance-mediating protease mutations is demonstrated herein by two exemplary protease PACE experiments, that evolve and analyze hepatitis C virus (HCV) protease variants in the presence of danoprevir or asunaprevir, two hepatitis C virus (HCV) protease inhibitor drug candidates in clinical trials. The evolved HCV protease variants exhibited up to 30-fold drug resistance after only 1 to 3 days of PACE. Predominant mutations in resistant variants were identified, and these predominant mutations matched those observed to arise in human patients treated with danoprevir or asunaprevir, demonstrating that PACE of a protease can rapidly identify the vulnerabilities of drug candidates to the evolution of clinically relevant drug resistance.

[0013] Some aspects of this disclosure provide methods for protease evolution. In some embodiments, such methods comprise providing a phage encoding a protease of interest, contacting a population of host cells with the phage, incubating the phage in a flow of host cells under circumstances that allow for the phage to mutate, replicate in the host cells, and infect fresh host cells. Typically, the host cells harbor an engineered expression system that links expression of a gene required for the generation of infectious phage particles (*e.g.*, pIII) to a desired protease activity to be evolved, and thus confers a selective advantage to those phage encoding mutated variants of the protease that exhibit the desired activity in the pool of replicating and mutating phages, resulting in an enrichment of desired mutations over time, as such mutations replicate more efficiently in and escape dilution and washout from the flow of host cells. The methods provided herein may also include a step of

isolating a replicated vector from the host cell population at the end of the PACE experiment, encoding a mutated version of a protease having a desired activity.

[0014] The linkage of a desired protease activity to a selective replication advantage of phage encoding protease variants exhibiting such desired activity is typically provided by an expression system in the host cells in which the expression level of a gene required for the generation of infectious phage depends on the desired protease activity. This can be achieved, for example, by providing a gene required for the generation of infectious phage particles in the host cells under the control of a conditional promoter, and additionally providing a transcriptional activator in the host cells that can drive transcription from the conditional promoter. In some embodiments, the transcriptional activator is provided in an inactive form that is converted into an active form by the desirable protease activity. For example, in some embodiments, a transcriptional activator, such as an RNA polymerase, is provided as a fusion protein with a transcriptional inhibitor, for example, an RNA-pol lysozyme, wherein the activator and inhibitor domains are connected by a linker comprising a protease cleavage site. Upon cleavage of the linker by an evolved protease, the inhibition of the transcriptional activator is released, and the transcriptional activator can drive expression of the gene for the generation of infectious phage particles.

[0015] Some aspects of this disclosure provide methods for generating protease variants that are resistant to a protease inhibitor. Such methods are similar in general to the protease-evolution methods described herein, but protease evolution is typically performed in the presence of a protease inhibitor. The methods may include a step of isolating mutated variants of a protease at the end of the PACE experiment and analyzing mutations that confer resistance to the protease inhibitor. In some embodiments, the methods comprise analyzing a plurality of resistant protease variants and determining common or dominant mutations amongst such variants to determine which mutations can be causally linked to resistance to the inhibitor. In some embodiments, the methods comprise comparing the evolution of inhibitor-resistant protease variants using a plurality of candidate inhibitors. For example, in some embodiments, the evolution of inhibitor-resistant protease variants is determined in several parallel experiments using the same protease but different protease inhibitors. In some embodiments, the methods include selecting an inhibitor from a plurality of inhibitors based on a PACE experiment yielding no resistant protease variants to the inhibitor or based on the selected inhibitor having the highest number of mutations required for a target protease to gain resistance.

[0016] Some aspects of this disclosure provide fusion proteins that link protease activity to transcriptional activation. Typically, such fusion proteins comprise a transcriptional activator in an inactive form, for example, fused to an inhibitor of the transcriptional activator via a linker comprising a protease cleavage site. In some embodiments, proteolytic cleavage of the linker results in release of the inhibitor from the transcriptional activator and thus for a release of the inhibition of the activator. Some aspects of this disclosure provide nucleic acid constructs encoding a fusion protein as provided herein.

[0017] Some aspects of this disclosure provide mutagenesis plasmids that enhance the mutagenesis rate in the host cells during a PACE experiment. In some embodiments, the mutagenesis plasmids comprise a gene expression cassette encoding a component of *E. coli* translesion synthesis polymerase V, a deoxyadenosine methylase, and/or a hemimethylated-GATC binding domain, or any combination thereof. In some embodiments, the component of *E. coli* translesion synthesis polymerase V is *umuC*. In some embodiments, the deoxyadenosine methylase is *dam*. In some embodiments, the hemimethylated-GATC binding domain is *seqA*.

[0018] Some aspects of this disclosure provide kits comprising reagents and materials useful for the use of PACE for protease evolution. In some embodiments, the kit comprises (a) a phage vector encoding a phage backbone, *e.g.*, an M13 phage backbone, and a multiple cloning site for insertion of a nucleic acid sequence encoding a protease. In some embodiments, the vector or a replication product thereof can be packaged into infectious phage particles in the presence of other phage functions by suitable host cells, but lacks at least one gene required for the generation of infectious particles. In some embodiments, the kit comprises (b) an accessory plasmid comprising a nucleic acid sequence encoding the at least one gene of interest under the control of a promoter that is activated by a transcriptional activator. In some embodiments, the kit comprises (c) an expression construct encoding a fusion protein of the transcriptional activator that activates the promoter of (b) fused to an inhibitor of the transcriptional activator via a linker, and a multiple cloning site for insertion of a nucleic acid sequence encoding a protease cleavage site. In some embodiments, the kit further comprises a helper phage providing all phage functions except for the at least one gene required for the generation of infectious phage particles provided by the accessory plasmid of (b). In some embodiments, the kit comprises suitable host cells. In some embodiments, the kit further comprises a mutagenesis plasmid.

[0019] Some aspects of this disclosure provide a system for protease evolution, comprising an apparatus or bioreactor through which a flow of host cells can be directed at a flow rate that results in an average time of the host cells remaining in the lagoon that is sufficient for a phage vector to replicate, but not sufficiently long for the host cell to divide and proliferate, wherein the host cells comprise an expression system linking the expression level of a gene required for the generation of infectious phage to a desired protease activity as described herein. In some embodiments, the host cells in the apparatus comprise a gene required for the generation of infectious phage particles under the control of a conditional promoter, for example, on an accessory plasmid. In some embodiment, the host cells further comprise a transcriptional activator that can drive transcription from the conditional promoter, and in some such embodiments, the transcriptional activator is provided in an inactive form that is converted into an active form by the desirable protease activity.

[0020] The summary above is meant to illustrate and outline, in a non-limiting manner, some of the embodiments, advantages, features, and uses of the technology disclosed herein. The disclosure is, however, not limited to the embodiments described in the summary above. Other embodiments, advantages, features, and uses of the technology disclosed herein will be apparent from the Detailed Description, the Drawings, the Examples, and the Claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] **Figures 1A-D. Development of a system to link protease activity to gene expression.** (**A**) Protease-activated RNA polymerase (PA-RNAP). T7 RNAP is fused to the natural inhibitor T7 lysozyme through a linker containing a protease target substrate sequence. While the linker is intact, the complex preferentially adopts the lysozyme-bound, RNAP-inactive state. Proteolysis of the target sequence favors dissociation of the complex, freeing active T7 RNAP to transcribe genes downstream of the T7 promoter. This example used an accessory plasmid (AP) in which the T7 promoter drives a tandem gIII-luciferase (lux) cassette. (**B**) Sequences of the protein linkers containing a target protease substrate used for each PA-RNAP, with T7 lysozyme residues in blue, protease substrates in red, T7 RNAP residues in green, and linker regions in black. (**C**) Plasmids used for protease PACE. An accessory plasmid (AP) that has gIII and luciferase (lux) under the control of the T7 promoter serves as the source of gIII in the cells. A complementary plasmid (CP) constitutively expresses a PA-RNAP variant with a protease target substrate sequence embedded in the linker. (**D**) PA-RNAP gene expression response in *E. coli* cells. Host cells were transformed with (*i*) an AP containing the T7 promoter driving gIII-lux; (*ii*) a CP that constitutively

expresses a PA-RNAP including the TEV protease substrate, the HCV protease substrate, or the HRV protease substrate; and (*iii*) a plasmid that expresses TEV protease (orange bars), HCV protease (purple bars), or HRV protease (gray bars). Gene expression is activated only when the expressed protease cleaves the amino acid sequence on the PA-RNAP sensor. The luminescence experiment was performed in triplicate with error bars indicating the standard deviation.

[0022] Figures 2A-D. PA-RNAPs link protease activity to phage propagation. (**A**) The protease PACE system. Fixed volume vessels (lagoons) contain phage in which gIII is replaced with a gene encoding an evolving protease. The lagoon is fed with host cells that contain an AP with the T7 promoter driving gIII and a CP that expresses a PA-RNAP. Phage infect incoming cells and inject their genome containing a protease variant. Only if the protease variant can activate the PA-RNAP by cleaving the linker encoding the target protease substrate, gIII is expressed and that SP can propagate. (**B-D**) Enrichment of active proteases from mixed populations using PACE. At time 0, a lagoon was seeded with a 1,000-fold excess of non-cognate protease-encoding phage over cognate protease-encoding phage. The lagoon was continuously diluted with host cells containing a PA-RNAP with either the HCV (**B**), TEV (**C**), or HRV (**D**) protease substrates. Lagoon samples were periodically analyzed by PCR. In all three cases, phage encoding the cognate protease were rapidly enriched in the lagoon while phage encoding the non-cognate protease were depleted.

[0023] Figures 3A-B. HCV PA-RNAP response to protease inhibitors in *E. coli* cells. Host cells expressing the HCV PA-RNAP were incubated with the HCV protease inhibitors danoprevir (**A**) or asunaprevir (**B**) for 90 min, followed by inoculation with HCV protease encoding phage. After 3 hours, luminescence assays were used to quantify relative gene activation resulting from the PA-RNAP. Luminescence experiments were performed in triplicate with error bars depicting the standard deviation.

[0024] Figures 4A-E. Continuous evolution of drug resistance in HCV protease. (**A**, **B**) PACE condition timeline for evolution in the presence of danoprevir (**A**) or asunaprevir (**B**). The blue arrows indicate arabinose-induced enhanced mutagenesis, and the red arrow shows the timing and dosing of HCV protease inhibitors. PACE condition timeline for evolution in the presence of asunaprevir. (**C**) High-throughput sequencing data from phage populations in replicate lagoons (L1 and L2) subjected to danoprevir treatment at 28 h, asunaprevir treatment at 75 h, and no drug at 72 h. All mutations with frequencies more than 1% above the allele-specific error rate are shown. (**D**) *In vitro* analysis of danoprevir inhibition of mutant HCV proteases that evolved during PACE. (**E**) *In vitro* analysis of

asunaprevir inhibition of mutant HCV proteases that evolved during PACE. For (**D**) and (**E**), evolved HCV protease variants were expressed and purified, then assayed using an internally quenched fluorescent-substrate (Anaspec). *In vitro* analyses were performed in triplicate with error bars calculated as the standard deviation.

[0025] **Figure 5. Crystal structure of T7 lysozyme bound to T7 RNAP**. Generated from PDB-1ARO[32].

[0026] **Figure 6.** Vector map of complementary plasmid (CP) used in Example 1.

[0027] **Figure 7.** Vector map of expression plasmid (EP) used in Example 1.

[0028] **Figure 8.** Vector map of accessory plasmid (AP) used in Example 1.

[0029] **Figure 9.** Vector map of selection phage (SP) used in Example 1.

[0030] **Figure 10.** Vector map of mutagenesis plasmid (MP) used in Example 1.

[0031] **Figure 11. Western blot showing PA-RNAP is cleaved only by the protease that is known to recognize the target sequence.** Band sizes of ~120 kDa and ~100 kDa correspond to the full PA-RNAP construct and the cleaved RNAP, respectively.

[0032] **Figure 12**. Directed evolution of a reprogrammed HRV protease.

[0033] **Figure 13.** Directed evolution of a reprogrammed TEV protease.

[0034] **Figure 14.** Directed evolution of TEV variants with improved specificity using negative selection against undesired cleavage target sequences.

[0035] **Figure 15.** Directed evolution of TEV variants cleaving single-mutant substrate sequences.

[0036] **Figure 16.** Directed evolution of TEV variants cleaving single-mutant substrate sequences.

[0037] **Figure 17.** Directed evolution of TEV variants cleaving double-mutant substrate sequences.

[0038] **Figure 18.** Directed evolution of TEV variants cleaving triple-mutant substrate sequences.

[0039] **Figure 19.** Directed evolution of TEV variants cleaving a target IL23a peptide.

[0040] **Figure 20.** Cleavage efficiency of wild-type TEV protease on various substrate sequences.

[0041] **Figure 21**. Cleavage efficiency of two evolved TEV proteases on various substrates sequences.

[0042] **Figure 22.** Directed evolution of HCV protease variants that cleave macaque MAVS.

[0043] **Figure 23.** Directed evolution of HCV protease variants that cleave macaque MAVS.

DEFINITIONS

[0044] The term "accessory plasmid," as used herein, refers to a plasmid comprising a gene required for the generation of infectious viral particles under the control of a conditional promoter. In the context of continuous evolution of proteases described herein, transcription from the conditional promoter of the accessory plasmid is typically activated by a function of the protease to be evolved. Accordingly, the accessory plasmid serves the function of conveying a competitive advantage to those viral vectors in a given population of viral vectors that carry a gene of interest able to activate the conditional promoter. Only viral vectors carrying an "activating" version of the protease of interest will be able to induce expression of the gene required to generate infectious viral particles in the host cell, and, thus, allow for packaging and propagation of the viral genome in the flow of host cells. Vectors carrying non-activating versions of the protease of interest, on the other hand, will not induce expression of the gene required to generate infectious viral vectors, and, thus, will not be packaged into viral particles that can infect fresh host cells.

[0045] The term "cellstat," as used herein, refers to a culture vessel comprising host cells, in which the number of cells is substantially constant over time.

[0046] The term "continuous evolution," as used herein, refers to an evolution process, in which a population of nucleic acids encoding a protease of interest is subjected to multiple rounds of (a) replication, (b) mutation, and (c) selection to produce a desired evolved protease that is different from the original protease of interest, for example, in that it cuts a target site not cut by the original protease, cuts its target site in the presence of an inhibitor of the original protease, or cuts its target site at an increased efficiency. The multiple rounds can be performed without investigator intervention, and the steps (a)-(c) can be carried out simultaneously. Typically, the evolution procedure is carried out *in vitro*, for example, using cells in culture as host cells. In general, a continuous evolution process provided herein relies on a system in which a gene encoding a protease of interest is provided in a nucleic acid vector that undergoes a life-cycle including replication in a host cell and transfer to another host cell, wherein a critical component of the life-cycle is deactivated and reactivation of the component is dependent upon an activity of the protease of interest that is a result of a mutation in the nucleic acid vector.

[0047] The term "flow", as used herein in the context of host cells, refers to a stream of host cells, wherein fresh host cells not harboring the transfer vector (*e.g.*, the viral vector encoding the protease of interest) are being introduced into a host cell population, for example, a host cell population in a lagoon, remain within the population for a limited time, and are then removed from the host cell population. In a simple form, a host cell flow may be a flow through a tube, or a channel, for example, at a controlled rate. In other embodiments, a flow of host cells is directed through a lagoon that holds a volume of cell culture media and comprises an inflow and an outflow. The introduction of fresh host cells may be continuous or intermittent and removal may be passive, *e.g.*, by overflow, or active, *e.g.*, by active siphoning or pumping. Removal further may be random, for example, if a stirred suspension culture of host cells is provided, removed liquid culture media will contain freshly introduced host cells as well as cells that have been a member of the host cell population within the lagoon for some time. Even though, in theory, a cell could escape removal from the lagoon indefinitely, the average host cell will remain only for a limited period of time within the lagoon, which is determined mainly by the flow rate of the culture media (and suspended cells) through the lagoon. Since the viral vectors replicate in a flow of host cells, in which fresh, uninfected host cells are provided while infected cells are removed, multiple consecutive viral life cycles can occur without investigator interaction, which allows for the accumulation of multiple advantageous mutations in a single evolution experiment.

[0048] The term "fresh," as used herein in the context of host cells, and used interchangeably with the terms "non-infected" or "uninfected" in the context of host cells of viral vectors, refers to a host cell that does not harbor the vector or, in the context of viral vectors, has not been infected by the viral vector comprising a gene encoding a protease of interest as used in a continuous evolution process provided herein. A fresh host cell can, however, have been infected by a viral vector unrelated to the vector to be evolved or by a vector of the same or a similar type but not carrying the gene of interest.

[0049] The term "fusion protein" as used herein refers to a hybrid polypeptide which comprises protein sequences or domains from at least two different proteins. One protein may be located at the amino-terminal (N-terminal) portion of the fusion protein or at the carboxy-terminal (C-terminal) protein thus forming an "amino-terminal fusion protein" or a "carboxy-terminal fusion protein," respectively. In the context of protease evolution using a transcriptional activator fused to an inhibitor, a fusion protein typically comprises a transcriptional activator domain that can bind to and drive gene expression from the conditional promoter of the accessory plasmid. Such a transcriptional activator domain may

be able to drive gene expression by itself (*e.g.*, an RNA polymerase domain) or may recruit transcriptional machinery to the promoter (*e.g.*, a transcription factor domain). The fusion proteins provided herein may be produced by any method known in the art. For example, the proteins provided herein may be produced via recombinant methods. Methods for recombinant protein generation and expression are well known and include those described by Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)), the entire contents of which are incorporated herein by reference.

[0050] The term "gene of interest" or "gene encoding a protease of interest," as used herein, refers to a nucleic acid construct comprising a nucleotide sequence encoding a gene product, *e.g.*, a protease, of interest to be evolved in a continuous evolution process as provided herein. The term includes any variations of a gene of interest that are the result of a continuous evolution process according to methods provided herein. For example, in some embodiments, a gene of interest is a nucleic acid construct comprising a nucleotide sequence encoding a protease to be evolved, cloned into a viral vector, for example, a phage genome, so that the expression of the encoding sequence is under the control of one or more promoters in the viral genome. In other embodiments, a gene of interest is a nucleic acid construct comprising a nucleotide sequence encoding a protease to be evolved and a promoter operably linked to the encoding sequence. When cloned into a viral vector, for example, a phage genome, the expression of the encoding sequence of such genes of interest is under the control of the heterologous promoter and, in some embodiments, may also be influenced by one or more promoters comprised in the viral genome.

[0051] The term "helper phage," as used herein interchangeable with the terms "helper phagemid" and "helper plasmid," refers to a nucleic acid construct comprising a phage gene required for the phage life cycle, or a plurality of such genes, but lacking a structural element required for genome packaging into a phage particle. For example, a helper phage may provide a wild-type phage genome lacking a phage origin of replication. In some embodiments, a helper phage is provided that comprises a gene required for the generation of phage particles, but lacks a gene required for the generation of infectious particles, for example, a full-length pIII gene. In some embodiments, the helper phage provides only some, but not all, genes for the generation of infectious phage particles. Helper phages are useful to allow modified phages that lack a gene for the generation of infectious phage particles to complete the phage life cycle in a host cell. Typically, a helper phage will comprise the genes for the generation of infectious phage particles that are lacking in the

phage genome, thus complementing the phage genome. In the continuous evolution context, the helper phage typically complements the selection phage, but both lack a phage gene required for the production of infectious phage particles.

[0052] The terms "high copy number plasmid" and "low copy number plasmid" are art-recognized, and those of skill in the art will be able to ascertain whether a given plasmid is a high or low copy number plasmid. In some embodiments, a low copy number accessory plasmid is a plasmid exhibiting an average copy number of plasmid per host cell in a host cell population of about 5 to about 100. In some embodiments, a very low copy number accessory plasmid is a plasmid exhibiting an average copy number of plasmid per host cell in a host cell population of about 1 to about 10. In some embodiments, a very low copy number accessory plasmid is a single-copy per cell plasmid. In some embodiments, a high copy number accessory plasmid is a plasmid exhibiting an average copy number of plasmid per host cell in a host cell population of about 100 to about 5000.

[0053] The term "host cell," as used herein, refers to a cell that can host, replicate, and transfer a phage vector useful for a continuous evolution process as provided herein. In embodiments where the vector is a viral vector, a suitable host cell is a cell that can be infected by the viral vector, can replicate it, and can package it into viral particles that can infect fresh host cells. A cell can host a viral vector if it supports expression of genes of viral vector, replication of the viral genome, and/or the generation of viral particles. One criterion to determine whether a cell is a suitable host cell for a given viral vector is to determine whether the cell can support the viral life cycle of a wild-type viral genome that the viral vector is derived from. For example, if the viral vector is a modified M13 phage genome, as provided in some embodiments described herein, then a suitable host cell would be any cell that can support the wild-type M13 phage life cycle. Suitable host cells for viral vectors useful in continuous evolution processes are well known to those of skill in the art, and the disclosure is not limited in this respect.

[0054] The term "infectious viral particle," as used herein, refers to a viral particle able to transport the viral genome it comprises into a suitable host cell. Not all viral particles are able to transfer the viral genome to a suitable host cell. Particles unable to accomplish this are referred to as non-infectious viral particles. In some embodiments, a viral particle comprises a plurality of different coat proteins, wherein one or some of the coat proteins can be omitted without compromising the structure of the viral particle. In some embodiments, a viral particle is provided in which at least one coat protein cannot be omitted without the loss of infectivity. If a viral particle lacks a protein that confers infectivity, the viral particle is not

infectious. For example, an M13 phage particle that comprises a phage genome packaged in a coat of phage proteins (*e.g.*, pVIII) but lacks pIII (protein III) is a non-infectious M13 phage particle because pIII is essential for the infectious properties of M13 phage particles.

[0055] The term "lagoon," as used herein, refers to a culture vessel or bioreactor through which a flow of host cells is directed. When used for a continuous evolution process as provided herein, a lagoon typically holds a population of host cells and a population of viral vectors replicating within the host cell population, wherein the lagoon comprises an outflow through which host cells are removed from the lagoon and an inflow through which fresh host cells are introduced into the lagoon, thus replenishing the host cell population.

[0056] The term "mutagen," as used herein, refers to an agent that induces mutations or increases the rate of mutation in a given biological system, for example, a host cell, to a level above the naturally occurring level of mutation in that system. Some exemplary mutagens useful for continuous evolution procedures are provided elsewhere herein, and other useful mutagens will be evident to those of skill in the art. Useful mutagens include, but are not limited to, ionizing radiation, ultraviolet radiation, base analogs, deaminating agents (*e.g.,* nitrous acid), intercalating agents (*e.g.*, ethidium bromide), alkylating agents (*e.g.,* ethylnitrosourea), transposons, bromine, azide salts, psoralen, benzene,3- Chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone (MX) (CAS no. 77439-76-0), O,O-dimethyl-S-(phthalimidomethyl)phosphorodithioate (phos-met) (CAS no. 732-11- 6), formaldehyde (CAS no. 50-00-0), 2-(2-furyl)-3-(5-nitro-2-furyl)acrylamide (AF-2) (CAS no. 3688-53-7), glyoxal (CAS no. 107-22-2), 6-mercaptopurine (CAS no. 50-44- 2), N-(trichloromethylthio)-4-cyclohexane-l,2-dicarboximide (captan) (CAS no. 133- 06-2), 2-aminopurine (CAS no. 452-06-2), methyl methane sulfonate (MMS) (CAS No. 66-27-3), 4-nitroquinoline 1-oxide (4-NQO) (CAS No. 56-57-5), N4-Aminocytidine (CAS no. 57294-74-3), sodium azide (CAS no. 26628-22-8), N-ethyl-N-nitrosourea (ENU) (CAS no. 759-73-9), N-methyl-N-nitrosourea (MNU) (CAS no. 820-60-0), 5- azacytidine (CAS no. 320-67-2), cumene hydroperoxide (CHP) (CAS no. 80-15-9), ethyl methanesulfonate (EMS) (CAS no. 62-50-0), N-ethyl-N -nitro-N-nitrosoguanidine (ENNG) (CAS no. 4245-77-6), N-methyl-N-nitro-N-nitrosoguanidine (MNNG) (CAS no. 70-25-7), 5-diazouracil (CAS no. 2435-76-9), and *t*-butyl hydroperoxide (BHP) (CAS no. 75-91-2). Additional mutagens can be used in continuous evolution procedures as provided herein, and the invention is not limited in this respect.

[0057] The term "mutagenesis plasmid," as used herein, refers to a plasmid comprising a nucleic acid sequence encoding a gene product or a combination of gene

products that act(s) as a mutagen. In some embodiments, a mutagenesis plasmid may encode a DNA polymerase lacking a proofreading capability. In some embodiments, the mutagenesis plasmid may encode a gene product involved in the bacterial SOS stress response, for example, a component of a bacterial translesion synthesis polymerase V. In some embodiments, the mutagenesis plasmid may encode a deoxyadenosine methylase. In some embodiments, the mutagenesis plasmid may encode a hemimethylated-GATC binding domain. In some non-limiting embodiments, the mutagenesis plasmid encodes *UmuC* (a component of E. coli translesion synthesis polymerase V), *dam* (deoxyadenosine methylase), and/or *seqA* (hemimethylated-GATC binding domain), or any combination thereof.

[0058] The term "nucleic acid," as used herein, refers to a polymer of nucleotides. The polymer may include natural nucleosides (*i.e.*, adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine), nucleoside analogs (*e.g.,* 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 3-methyl adenosine, 5-methylcytidine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 7-deazaadenosine, 7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, 4-acetylcytidine, 5-(carboxyhydroxymethyl)uridine, dihydrouridine, methylpseudouridine, 1-methyl adenosine, 1-methyl guanosine, N6-methyl adenosine, and 2-thiocytidine), chemically modified bases, biologically modified bases (*e.g.*, methylated bases), intercalated bases, modified sugars (*e.g.*, 2′-fluororibose, ribose, 2′-deoxyribose, 2′-O-methylcytidine, arabinose, and hexose), or modified phosphate groups (*e.g.*, phosphorothioates and 5' -N-phosphoramidite linkages).

[0059] The term "phage," as used herein interchangeably with the term "bacteriophage," refers to a virus that infects bacterial cells. Typically, phages consist of an outer protein capsid enclosing genetic material. The genetic material can be ssRNA, dsRNA, ssDNA, or dsDNA, in either linear or circular form. Phages and phage vectors are well known to those of skill in the art and non-limiting examples of phages that are useful for carrying out the methods provided herein are λ (Lysogen), T2, T4, T7, T12, R17, M13, MS2, G4, P1, P2, P4, Phi X174, N4, Φ6, and Φ29. In certain embodiments, the phage utilized in the present invention is M13. Additional suitable phages and host cells will be apparent to those of skill in the art and the invention is not limited in this aspect. For an exemplary description of additional suitable phages and host cells, see Elizabeth Kutter and Alexander Sulakvelidze: *Bacteriophages: Biology and Applications*. CRC Press; 1st edition (December 2004), ISBN: 0849313368; Martha R. J. Clokie and Andrew M. Kropinski: *Bacteriophages:*

Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions (Methods in Molecular Biology) Humana Press; 1st edition (December, 2008), ISBN: 1588296822; Martha R. J. Clokie and Andrew M. Kropinski: *Bacteriophages: Methods and Protocols, Volume 2: Molecular and Applied Aspects (Methods in Molecular Biology)* Humana Press; 1st edition (December 2008), ISBN: 1603275649; all of which are incorporated herein in their entirety by reference for disclosure of suitable phages and host cells as well as methods and protocols for isolation, culture, and manipulation of such phages).

[0060] The term "phage-assisted continuous evolution (PACE)," as used herein, refers to continuous evolution that employs phage as viral vectors. PACE technology has been described previously, for example, in International PCT Application, PCT/US2009/056194, filed September 8, 2009, published as WO 2010/028347 on March 11, 2010; International PCT Application, PCT/US2011/066747, filed December 22, 2011, published as WO 2012/088381 on June 28, 2012; and U.S. Application, U.S.S.N. 13/922,812, filed June 20, 2013, each of which is incorporated herein by reference.

[0061] The term "promoter" is art-recognized and refers to a nucleic acid molecule with a sequence recognized by the cellular transcription machinery and able to initiate transcription of a downstream gene. A promoter can be constitutively active, meaning that the promoter is always active in a given cellular context, or conditionally active, meaning that the promoter is only active in the presence of a specific condition. For example, a conditional promoter may only be active in the presence of a specific protein that connects a protein associated with a regulatory element in the promoter to the basic transcriptional machinery, or only in the absence of an inhibitory molecule. A subclass of conditionally active promoters are inducible promoters that require the presence of a small molecule "inducer" for activity. Examples of inducible promoters include, but are not limited to, arabinose-inducible promoters, Tet-on promoters, and tamoxifen-inducible promoters. A variety of constitutive, conditional, and inducible promoters are well known to the skilled artisan, and the skilled artisan will be able to ascertain a variety of such promoters useful in carrying out the instant invention, which is not limited in this respect.

[0062] The term "protease inhibitor," as used herein, refers to a molecule that inhibits the activity of a protease. Many naturally occurring protease inhibitors are known to those of skill in the art. The term also embraces non-naturally occurring protease inhibitors, including, but not limited to, small molecule protease inhibitors. Suitable protease inhibitors will be apparent to those of skill in the art and include, without limitation, protease inhibitors listed in the MEROPS database, accessible at merops.sanger.ac.uk and described in

Rawlings *et al.*, (2014) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* 42, D503-D509, the entire contents of each of which are incorporated herein by reference.

[0063] The term "protease," as used herein, refers to an enzyme that catalyzes the hydrolysis of a peptide bond linking amino acid residues together within a protein. The term embraces both naturally occurring and engineered proteases. Many proteases are known in the art. Proteases can be classified by their catalytic residue, and protease classes include, without limitation, serine proteases (serine alcohol), threonine proteases (threonine secondary alcohol), cysteine proteases (cysteine thiol), aspartate proteases (aspartate carboxylic acid), glutamic acid proteases (glutamate carboxylic acid), and metalloproteases (metal ion, *e.g.*, zinc). The structures in parentheses correlate to the respective catalytic moiety of proteases of each class. Some proteases are highly promiscuous and cleave a wide range of protein substrates, *e.g.*, trypsin or pepsin. Other proteases are highly specific and only cleave substrates with a specific sequence. Some blood clotting proteases such as, for example, thrombin, and some viral proteases such as, for example, HCV or TEV protease, are highly specific proteases. Proteases that cleave in a very specific manner typically bind to multiple amino acid residues of their substrate. Suitable proteases and protease cleavage sites, also sometimes referred to as "protease substrates," will be apparent to those of skill in the art and include, without limitation, proteases listed in the MEROPS database, accessible at merops.sanger.ac.uk and described in Rawlings *et al.*, (2014) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* 42, D503-D509, the entire contents of each of which are incorporated herein by reference. The disclosure is not limited in this respect.

[0064] The term "protein," as used herein refers to a polymer of amino acid residues linked together by peptide bonds. The term, as used herein, refers to proteins, polypeptides, and peptide of any size, structure, or function. Typically, a protein will be at least three amino acids long. A protein may refer to an individual protein or a collection of proteins. Inventive proteins preferably contain only natural amino acids, although non-natural amino acids (*i.e.*, compounds that do not occur in nature but that can be incorporated into a polypeptide chain; see, for example, cco.caltech.edu/~dadgrp/Unnatstruct.gif, which displays structures of non-natural amino acids that have been successfully incorporated into functional ion channels) and/or amino acid analogs as are known in the art may alternatively be employed. Also, one or more of the amino acids in an inventive protein may be modified, for example, by the addition of a chemical entity such as a carbohydrate group, a hydroxyl group,

a phosphate group, a farnesyl group, an isofarnesyl group, a fatty acid group, a linker for conjugation, functionalization, or other modification, *etc.* A protein may also be a single molecule or may be a multi-molecular complex. A protein may be just a fragment of a naturally occurring protein or peptide. A protein may be naturally occurring, recombinant, or synthetic, or any combination of these.

[0065] The term "replication product," as used herein, refers to a nucleic acid that is the result of viral genome replication by a host cell. This includes any viral genomes synthesized by the host cell from a viral genome inserted into the host cell. The term includes non-mutated as well as mutated replication products.

[0066] The term "selection phage," as used herein interchangeably with the term "selection plasmid," refers to a modified phage that comprises a gene of interest to be evolved and lacks a full-length gene encoding a protein required for the generation of infectious phage particles. For example, some M13 selection phages provided herein comprise a nucleic acid sequence encoding a protease to be evolved, *e.g.*, under the control of an M13 promoter, and lack all or part of a phage gene encoding a protein required for the generation of infectious phage particles, *e.g.*, gI, gII, gIII, gIV, gV, gVI, gVII, gVIII, gIX, or gX, or any combination thereof. For example, some M13 selection phages provided herein comprise a nucleic acid sequence encoding a protease to be evolved, *e.g.*, under the control of an M13 promoter, and lack all or part of a gene encoding a protein required for the generation of infective phage particles, *e.g.*, the gIII gene encoding the pIII protein.

[0067] The terms "small molecule" and "organic compound" are used interchangeably herein and refer to molecules, whether naturally-occurring or artificially created (*e.g.*, via chemical synthesis) that have a relatively low molecular weight. Typically, an organic compound contains carbon. An organic compound may contain multiple carbon-carbon bonds, stereocenters, and other functional groups (*e.g.*, amines, hydroxyl, carbonyls, or heterocyclic rings). In some embodiments, organic compounds are monomeric and have a molecular weight of less than about 1500 g/mol. In certain embodiments, the molecular weight of the small molecule is less than about 1000 g/mol or less than about 500 g/mol. In certain embodiments, the small molecule is a therapeutic drug or drug candidate, for example, a drug or drug candidate that is in clinical or pre-clinical trials or that has already been deemed safe and effective for use in humans or animals by the appropriate governmental agency or regulatory body.

[0068] The term "turbidostat," as used herein, refers to a culture vessel comprising host cells in suspension culture, in which the turbidity of the culture medium is substantially

essentially constant over time. In some embodiments, the turbidity of a suspension culture, for example, of bacterial cells, is a measure for the cell density in the culture medium. In some embodiments, a turbidostat comprises an inflow of fresh media and an outflow, and a controller that regulates the flow into and/or out of the turbidostat based on the turbidity of the suspension culture in the turbidostat.

[0069] The term "vector," as used herein, refers to a nucleic acid that can be modified to encode a protease of interest and that is able to enter into a host cell, mutate and replicate within the host cell, and then transfer a replicated form of the vector into another host cell. Exemplary suitable vectors include viral vectors, such as retroviral vectors or bacteriophages, and conjugative plasmids. Additional suitable vectors will be apparent to those of skill in the art based on the instant disclosure.

[0070] The term "viral life cycle," as used herein, refers to the viral reproduction cycle comprising insertion of the viral genome into a host cell, replication of the viral genome in the host cell, and packaging of a replication product of the viral genome into a viral particle by the host cell.

[0071] The term "viral particle," as used herein, refers to a viral genome, for example, a DNA or RNA genome, that is associated with a coat of a viral protein or proteins, and, in some cases, with an envelope of lipids. For example, a phage particle comprises a phage genome packaged into a protein encoded by the wild type phage genome.

[0072] The term "viral vector," as used herein, refers to a nucleic acid comprising a viral genome that, when introduced into a suitable host cell, can be replicated and packaged into viral particles able to transfer the viral genome into another host cell. The term viral vector extends to vectors comprising truncated or partial viral genomes. For example, in some embodiments, a viral vector is provided that lacks a gene encoding a protein essential for the generation of infectious viral particles. In suitable host cells, for example, host cells comprising the lacking gene under the control of a conditional promoter, however, such truncated viral vectors can replicate and generate viral particles able to transfer the truncated viral genome into another host cell. In some embodiments, the viral vector is a phage, for example, a filamentous phage (*e.g.*, an M13 phage). In some embodiments, a viral vector, for example, a phage vector, is provided that comprises a gene encoding the protease of interest to be evolved.

DETAILED DESCRIPTION

Introduction

[0073] Among the more than 600 naturally occurring proteases that have been described[1] are enzymes that have proven to be important catalysts of industrial processes, tools for proteome analysis, and life-saving pharmaceuticals[2-5]. Recombinant human proteases including thrombin, factor VIIa, and tissue plasminogen activator are widely used drugs for the treatment of blood clotting diseases[4]. In addition, the potential of protease-based therapeutics to address disease in a manner analogous to that of antibody drugs,[6,7] but with catalytic turnover, has been recognized for several decades[4,8]. Natural proteases, however, typically target only a narrowly defined set of substrates, limiting their therapeutic potential. The directed evolution of proteases in principle could generate enzymes with tailor-made specificities, but laboratory-evolved proteases are frequently non-specific, weakly active, or only modestly altered in their substrate specificity, limiting their utility[9-14].

[0074] In addition to their importance as current and future therapeutic agents, proteases have also proven to be major drug targets for diseases including cardiovascular disease, infectious disease, and cancer[15,16]. While drug specificity and potency are characterized and optimized during pre-clinical studies, the evolution of drug resistance is often not well understood until it arises in patients, despite the strong relationship between drug resistance vulnerability and a lack of therapeutic efficacy.

[0075] For example, resistance to HIV and HCV protease inhibitors can arise in as few as two days of clinical use[17] and frequently leads to viral rebound and poor treatment outcomes[18-21]. The speed with which drug resistance can arise in the clinic endangers patients and puts years of drug development efforts prior to such a determination at risk. Characterizing the potential of protease inhibitors to be overcome by the evolution of drug resistance using methods such as mammalian cell culture, animal models, or yeast display-based laboratory evolution, is time- and labor-intensive[22,23]. As a result, identifying drug resistance vulnerabilities of early-stage preclinical candidates is not common practice.

[0076] Phage-assisted continuous evolution (PACE) can serve as a rapid, high-throughput method to evolve protease enzymes and to reveal resistance to protease inhibitor drug candidates, analogous to previous uses of stepwise protein evolution to study antibiotic resistance[24]. During PACE, continuously replicating viral vectors, e.g., bacteriophage vectors, in a fixed-volume vessel (a "lagoon") carry an evolving gene of interest. Phage with genes encoding proteins with the desired target activity preferentially replicate because target activity triggers the production of a gene required for the transfer of the vector from one cell

to another, *e.g.*, a gene required for the generation of infectious phage particles, such as, for example, pIII, an essential component in the bacteriophage life cycle[25]. Because the lagoon is continuously diluted by a constant influx of host cells, e.g., *E. coli* cells, phage vectors encoding inactive variants produce non-infectious progeny that are rapidly diluted out of the lagoon. Dilution occurs faster than cell division but slower than phage replication, ensuring that mutations only accumulate in the phage genome. Because evolution during PACE takes place continuously without researcher intervention, hundreds of rounds of evolution can be performed per day or per week.

[0077] Some aspects of this disclosure are based on the recognition that PACE can be employed for the directed evolution of proteases. Proteases may require many successive mutations to remodel complex networks of contacts with polypeptide substrates[26,27], and are thus not readily manipulated by conventional, iterative evolution methods. The ability of PACE to perform the equivalent of hundreds of rounds of iterative evolution methods within days enables complex protease evolution experiments, that are impractical with conventional methods. Some aspects of this disclosure are based on the recognition that the speed of PACE also enables the rapid identification of mutations that confer resistance to protease inhibitors, for example, to therapeutic protease inhibitors. Such insights translate into methods for selecting drug candidates based on a determination of the likelihood of a target protease targeted by the drug evolving into a resistant mutant, and informs the design of new protease inhibitors for therapeutic and research use.

[0078] This disclosure provides data illustrating the feasibility of PACE-mediated protease evolution for non-limiting examples of the continuous directed evolution of proteases. The exemplary systems described in the Examples section employ an engineered protease-activated RNA polymerase (PA-RNAP) to transduce protease-mediated polypeptide cleavage events into changes in gene expression that support phage propagation during PACE. The successful link of the phage lifecycle to protease activity using PA-RNAP was validated for three distinct proteases. Protease PACE was also performed in the presence of therapeutic protease inhibitors, for example, danoprevir and asunaprevir, two hepatitis C virus (HCV) protease inhibitor drug candidates currently in clinical trials. Under those conditions, protease PACE rapidly evolved HCV protease variants that are resistant to each drug candidate. The PACE-evolved HCV protease variants were observed to be dominated by mutations previously observed in patients treated with these drug candidates, indicating that protease PACE can be used to recapitulate or predict the emergence of resistant protease variants in a clinical context and thus to support the selection of drug candidates for clinical

development. Together, the proof-of-concept findings described herein establish a new platform to rapidly generate proteases with novel properties through continuous evolution, and to reveal the vulnerability of protease inhibitors to the evolution of drug resistance.

[0079] PACE technology has been described previously, for example, in International PCT Application, PCT/US2009/056194, filed September 8, 2009, published as WO 2010/028347 on March 11, 2010; International PCT Application, PCT/US2011/066747, filed December 22, 2011, published as WO 2012/088381 on June 28, 2012; and U.S. Application, U.S.S.N. 13/922,812, filed June 20, 2013, each of which is incorporated herein by reference. Those of skill in the art will understand that the PACE technology, strategies, methods, and reagents provided herein can be used in combination with many aspects of the PACE technology described in those applications, for example, with the apparatuses, lagoons, host cell types, cell flow parameters, negative selection strategies, *etc.*, disclosed in these applications.

Continuous Evolution Methods

[0080] Some aspects of this disclosure provide methods for evolution of a protease. In some embodiments, a method of evolution of a protease is provided that comprises (a) contacting a population of host cells with a population of vectors comprising a gene encoding a protease. The vectors are typically deficient in at least one gene required for the transfer of the phage vector from one cell to another, *e.g.*, a gene required for the generation of infectious phage particles. In some embodiments of the provided methods, (1) the host cells are amenable to transfer of the vector; (2) the vector allows for expression of the protease in the host cell, can be replicated by the host cell, and the replicated vector can transfer into a second host cell; and (3) the host cell expresses a gene product encoded by the at least one gene for the generation of infectious phage particles (a) in response to the activity of the protease, and the level of gene product expression depends on the activity of the protease. The methods of protease evolution provided herein typically comprise (b) incubating the population of host cells under conditions allowing for mutation of the gene encoding the protease, and the transfer of the vector comprising the gene encoding the protease of interest from host cell to host cell. The host cells are removed from the host cell population at a certain rate, *e.g.*, at a rate that results in an average time a host cell remains in the cell population that is shorter than the average time a host cell requires to divide, but long enough for the completion of a life cycle (uptake, replication, and transfer to another host cell) of the vector. The population of host cells is replenished with fresh host cells that do not harbor the

vector. In some embodiments, the rate of replenishment with fresh cells substantially matches the rate of removal of cells from the cell population, resulting in a substantially constant cell number or cell density within the cell population. The methods of protease evolution provided herein typically also comprise (c) isolating a replicated vector from the host cell population in (b), wherein the replicated vector comprises a mutated version of the gene encoding the protease.

[0081] In some embodiments, a gene encoding a protease of interest is transferred from host cell to host cell in a manner dependent on the activity of the protease of interest. In some embodiments, the transfer vector is a virus infecting and replicating in the host cells, for example, a bacteriophage or a retroviral vector. In some embodiments, the viral vector is a phage vector infecting bacterial host cells. In some embodiments, the transfer vector is a retroviral vector, for example, a lentiviral vector or a vesicular stomatitis virus vector, infecting human or mouse cells. In some embodiments, the transfer vector is a conjugative plasmid transferred from a donor bacterial cell to a recipient bacterial cell.

[0082] In some embodiments, the nucleic acid vector comprising the gene encoding a protease of interest is a phage, a viral vector, or naked DNA (*e.g.*, a mobilization plasmid). In some embodiments, transfer of the gene encoding the protease of interest from cell to cell is via infection, transfect ion, transduction, conjugation, or uptake of naked DNA, and efficiency of cell-to-cell transfer (*e.g.*, transfer rate) is dependent on an activity of the protease of interest or a mutated version thereof. For example, in some embodiments, the nucleic acid vector is a phage harboring the gene encoding a protease of interest, and the efficiency of phage transfer (via infection) is dependent on the activity of the protease of interest in that a protein for the generation of infectious phage particles (*e.g.*, pIII for M13 phage) is expressed in the host cells only in the presence of a desired protease activity, such as the cleavage of a specific protease cleavage site or the cleavage of a target site under specific circumstances, *e.g.*, in the presence of a protease inhibitor.

[0083] Some embodiments provide a continuous evolution system, in which a population of viral vectors, *e.g.*, M13 phage vectors, comprising a gene encoding a protease of interest to be evolved replicates in a flow of host cells, *e.g.*, a flow through a lagoon, wherein the viral vectors are deficient in a gene encoding a protein that is essential for the generation of infectious viral particles, and wherein that gene is in the host cell under the control of a conditional promoter the activity of which depends on the activity of the protease of interest. In some embodiments, transcription from the conditional promoter may be activated by cleavage of a fusion protein comprising a transcription factor and an inhibitory

protein fused to the transcriptional activator via a linker comprising a target site of the protease.

[0084] Viral vectors, in which the gene encoding the protease of interest has not acquired a mutation conferring the desired function, will not activate the conditional promoter, or only achieve minimal activation, while any mutation in the gene of interest that confers the desired mutation will result in activation of the conditional promoter. Since the conditional promoter controls an essential protein for the viral life cycle, activation of this promoter directly corresponds to an advantage in viral spread and replication for those vectors that have acquired an advantageous mutation.

[0085] In some embodiments, the viral vector provided is a phage In some embodiments, the phage is a filamentous phage. In some embodiments, the phage is an M13 phage. M13 phages are well known to those in the art and the biology of M13 phages has extensively been studied. A schematic representation of the wild-type M13 genome is provided in Figure 16. Wild type M13 phage particles comprise a circular, single-stranded genome of approximately 6.4kb. The wilt-type genome includes ten genes, gI-gX, which, in turn, encode the ten M13 proteins, pI-pX, respectively. gVIII encodes pVIII, also often referred to as the major structural protein of the phage particles, while gIII encodes pIII, also referred to as the minor coat protein, which is required for infectivity of M13 phage particles.

[0086] The M13 life cycle includes attachment of the phage to the sex pilus of a suitable bacterial host cell via the pIII protein and insertion of the phage genome into the host cell. The circular, single-stranded phage genome is then converted to a circular, double-stranded DNA, also termed the replicative form (RF), from which phage gene transcription is initiated. The wild type M13 genome comprises nine promoters and two transcriptional terminators as well as an origin of replication. This series of promoters provides a gradient of transcription such that the genes nearest the two transcriptional terminators (gVIII and IV) are transcribed at the highest levels. In wild-type M13 phage, transcription of all 10 genes proceeds in same direction. One of the phage-encode proteins, pII, initiates the generation of linear, single-stranded phage genomes in the host cells, which are subsequently circularized, and bound and stabilized by pV. The circularized, single-stranded M13 genomes are then bound by pVIII, while pV is stripped off the genome, which initiates the packaging process. At the end of the packaging process, multiple copies of pIII are attached to wild-type M13 particles, thus generating infectious phage ready to infect another host cell and concluding the life cycle.

[0087] The M13 phage genome can be manipulated, for example, by deleting one or more of the wild type genes, and/or inserting a heterologous nucleic acid construct into the genome. M13 does not have stringent genome size restrictions, and insertions of up to 42 kb have been reported. This allows M13 phage vectors to be used in continuous evolution experiments to evolve genes of interest without imposing a limitation on the length of the gene to be involved.

[0088] The M13 phage has been well characterized and the genomic sequence of M13 has been reported. Representative M13 genomic sequences can be retrieved from public databases, and an exemplary sequence is provided in entry V00604 of the National Center for Biotechnology Information (NCBI) database (www.ncbi.nlm.nih.gov):

[0089] Phage M13 genome:

>gi|56713234|emb|V00604.2| Phage M13 genome
AACGCTACTACTATTAGTAGAATTGATGCCACCTTTTCAGCTCGCGCCCCAAATGAAAATAT
AGCTAAACAGGTTATTGACCATTTGCGAAATGTATCTAATGGTCAAACTAAATCTACTCGTT
CGCAGAATTGGGAATCAACTGTTACATGGAATGAAACTTCCAGACACCGTACTTTAGTTGCA
TATTTAAAACATGTTGAGCTACAGCACCAGATTCAGCAATTAAGCTCTAAGCCATCCGCAAA
AATGACCTCTTATCAAAAGGAGCAATTAAAGGTACTCTCTAATCCTGACCTGTTGGAGTTTG
CTTCCGGTCTGGTTCGCTTTGAAGCTCGAATTAAAACGCGATATTTGAAGTCTTTCGGGCTT
CCTCTTAATCTTTTTGATGCAATCCGCTTTGCTTCTGACTATAATAGTCAGGGTAAAGACCT
GATTTTTGATTTATGGTCATTCTCGTTTTCTGAACTGTTTAAAGCATTTGAGGGGGATTCAA
TGAATATTTATGACGATTCCGCAGTATTGGACGCTATCCAGTCTAAACATTTTACTATTACC
CCCTCTGGCAAAACTTCTTTTGCAAAAGCCTCTCGCTATTTTGGTTTTTATCGTCGTCTGGT
AAACGAGGGTTATGATAGTGTTGCTCTTACTATGCCTCGTAATTCCTTTTGGCGTTATGTAT
CTGCATTAGTTGAATGTGGTATTCCTAAATCTCAACTGATGAATCTTTCTACCTGTAATAAT
GTTGTTCCGTTAGTTCGTTTTATTAACGTAGATTTTTCTTCCCAACGTCCTGACTGGTATAA
TGAGCCAGTTCTTAAAATCGCATAAGGTAATTCACAATGATTAAAGTTGAAATTAAACCATC
TCAAGCCCAATTTACTACTCGTTCTGGTGTTTCTCGTCAGGGCAAGCCTTATTCACTGAATG
AGCAGCTTTGTTACGTTGATTTGGGTAATGAATATCCGGTTCTTGTCAAGATTACTCTTGAT
GAAGGTCAGCCAGCCTATGCGCCTGGTCTGTACACCGTTCATCTGTCCTCTTTCAAAGTTGG
TCAGTTCGGTTCCCTTATGATTGACCGTCTGCGCCTCGTTCCGGCTAAGTAACATGGAGCAG
GTCGCGGATTTCGACACAATTTATCAGGCGATGATACAAATCTCCGTTGTACTTTGTTTCGC
GCTTGGTATAATCGCTGGGGGTCAAAGATGAGTGTTTTAGTGTATTCTTTCGCCTCTTTCGT
TTTAGGTTGGTGCCTTCGTAGTGGCATTACGTATTTTACCCGTTTAATGGAAACTTCCTCAT
GAAAAAGTCTTTAGTCCTCAAAGCCTCTGTAGCCGTTGCTACCCTCGTTCCGATGCTGTCTT
TCGCTGCTGAGGGTGACGATCCCGCAAAAGCGGCCTTTAACTCCCTGCAAGCCTCAGCGACC
GAATATATCGGTTATGCGTGGGCGATGGTTGTTGTCATTGTCGGCGCAACTATCGGTATCAA
GCTGTTTAAGAAATTCACCTCGAAAGCAAGCTGATAAACCGATACAATTAAAGGCTCCTTTT
GGAGCCTTTTTTTTTGGAGATTTTCAACATGAAAAAATTATTATTCGCAATTCCTTTAGTTG
TTCCTTTCTATTCTCACTCCGCTGAAACTGTTGAAAGTTGTTTAGCAAAACCCCATACAGAA
AATTCATTTACTAACGTCTGGAAAGACGACAAAACTTTAGATCGTTACGCTAACTATGAGGG
TTGTCTGTGGAATGCTACAGGCGTTGTAGTTTGTACTGGTGACGAAACTCAGTGTTACGGTA
CATGGGTTCCTATTGGGCTTGCTATCCCTGAAAATGAGGGTGGTGGCTCTGAGGGTGGCGGT
TCTGAGGGTGGCGGTTCTGAGGGTGGCGGTACTAAACCTCCTGAGTACGGTGATACACCTAT
TCCGGGCTATACTTATATCAACCCTCTCGACGGCACTTATCCGCCTGGTACTGAGCAAAACC
CCGCTAATCCTAATCCTTCTCTTGAGGAGTCTCAGCCTCTTAATACTTTCATGTTTCAGAAT

AATAGGTTCCGAAATAGGCAGGGGGCATTAACTGTTTATACGGGCACTGTTACTCAAGGCAC
TGACCCCGTTAAAACTTATTACCAGTACACTCCTGTATCATCAAAAGCCATGTATGACGCTT
ACTGGAACGGTAAATTCAGAGACTGCGCTTTCCATTCTGGCTTTAATGAGGATCCATTCGTT
TGTGAATATCAAGGCCAATCGTCTGACCTGCCTCAACCTCCTGTCAATGCTGGCGGCGGCTC
TGGTGGTGGTTCTGGTGGCGGCTCTGAGGGTGGTGGCTCTGAGGGTGGCGGTTCTGAGGGTG
GCGGCTCTGAGGGAGGCGGTTCCGGTGGTGGCTCTGGTTCCGGTGATTTTGATTATGAAAAG
ATGGCAAACGCTAATAAGGGGGCTATGACCGAAAATGCCGATGAAAACGCGCTACAGTCTGA
CGCTAAAGGCAAACTTGATTCTGTCGCTACTGATTACGGTGCTGCTATCGATGGTTTCATTG
GTGACGTTTCCGGCCTTGCTAATGGTAATGGTGCTACTGGTGATTTTGCTGGCTCTAATTCC
CAAATGGCTCAAGTCGGTGACGGTGATAATTCACCTTTAATGAATAATTTCCGTCAATATTT
ACCTTCCCTCCCTCAATCGGTTGAATGTCGCCCTTTTGTCTTTAGCGCTGGTAAACCATATG
AATTTTCTATTGATTGTGACAAAATAAACTTATTCCGTGGTGTCTTTGCGTTTCTTTTATAT
GTTGCCACCTTTATGTATGTATTTTCTACGTTTGCTAACATACTGCGTAATAAGGAGTCTTA
ATCATGCCAGTTCTTTTGGGTATTCCGTTATTATTGCGTTTCCTCGGTTTCCTTCTGGTAAC
TTTGTTCGGCTATCTGCTTACTTTTCTTAAAAAGGGCTTCGGTAAGATAGCTATTGCTATTT
CATTGTTTCTTGCTCTTATTATTGGGCTTAACTCAATTCTTGTGGGTTATCTCTCTGATATT
AGCGCTCAATTACCCTCTGACTTTGTTCAGGGTGTTCAGTTAATTCTCCCGTCTAATGCGCT
TCCCTGTTTTTATGTTATTCTCTCTGTAAAGGCTGCTATTTTCATTTTTGACGTTAAACAAA
AAATCGTTTCTTATTTGGATTGGGATAAATAATATGGCTGTTTATTTTGTAACTGGCAAATT
AGGCTCTGGAAAGACGCTCGTTAGCGTTGGTAAGATTCAGGATAAAATTGTAGCTGGGTGCA
AAATAGCAACTAATCTTGATTTAAGGCTTCAAAACCTCCCGCAAGTCGGGAGGTTCGCTAAA
ACGCCTCGCGTTCTTAGAATACCGGATAAGCCTTCTATATCTGATTTGCTTGCTATTGGGCG
CGGTAATGATTCCTACGATGAAAATAAAAACGGCTTGCTTGTTCTCGATGAGTGCGGTACTT
GGTTTAATACCCGTTCTTGGAATGATAAGGAAAGACAGCCGATTATTGATTGGTTTCTACAT
GCTCGTAAATTAGGATGGGATATTATTTTTCTTGTTCAGGACTTATCTATTGTTGATAAACA
GGCGCGTTCTGCATTAGCTGAACATGTTGTTTATTGTCGTCGTCTGGACAGAATTACTTTAC
CTTTTGTCGGTACTTTATATTCTCTTATTACTGGCTCGAAAATGCCTCTGCCTAAATTACAT
GTTGGCGTTGTTAAATATGGCGATTCTCAATTAAGCCCTACTGTTGAGCGTTGGCTTTATAC
TGGTAAGAATTTGTATAACGCATATGATACTAAACAGGCTTTTTCTAGTAATTATGATTCCG
GTGTTTATTCTTATTTAACGCCTTATTTATCACACGGTCGGTATTTCAAACCATTAAATTTA
GGTCAGAAGATGAAATTAACTAAAATATATTTGAAAAAGTTTTCTCGCGTTCTTTGTCTTGC
GATTGGATTTGCATCAGCATTTACATATAGTTATATAACCCAACCTAAGCCGGAGGTTAAAA
AGGTAGTCTCTCAGACCTATGATTTTGATAAATTCACTATTGACTCTTCTCAGCGTCTTAAT
CTAAGCTATCGCTATGTTTTCAAGGATTCTAAGGGAAAATTAATTAATAGCGACGATTTACA
GAAGCAAGGTTATTCACTCACATATATTGATTTATGTACTGTTTCCATTAAAAAAGGTAATT
CAAATGAAATTGTTAAATGTAATTAATTTTGTTTTCTTGATGTTTGTTTCATCATCTTCTTT
TGCTCAGGTAATTGAAATGAATAATTCGCCTCTGCGCGATTTTGTAACTTGGTATTCAAAGC
AATCAGGCGAATCCGTTATTGTTTCTCCCGATGTAAAAGGTACTGTTACTGTATATTCATCT
GACGTTAAACCTGAAAATCTACGCAATTTCTTTATTTCTGTTTTACGTGCTAATAATTTTGA
TATGGTTGGTTCAATTCCTTCCATAATTCAGAAGTATAATCCAAACAATCAGGATTATATTG
ATGAATTGCCATCATCTGATAATCAGGAATATGATGATAATTCCGCTCCTTCTGGTGGTTTC
TTTGTTCCGCAAAATGATAATGTTACTCAAACTTTTAAAATTAATAACGTTCGGGCAAAGGA
TTTAATACGAGTTGTCGAATTGTTTGTAAAGTCTAATACTTCTAAATCCTCAAATGTATTAT
CTATTGACGGCTCTAATCTATTAGTTGTTAGTGCACCTAAAGATATTTTAGATAACCTTCCT
CAATTCCTTTCTACTGTTGATTTGCCAACTGACCAGATATTGATTGAGGGTTTGATATTTGA
GGTTCAGCAAGGTGATGCTTTAGATTTTTCATTTGCTGCTGGCTCTCAGCGTGGCACTGTTG
CAGGCGGTGTTAATACTGACCGCCTCACCTCTGTTTTATCTTCTGCTGGTGGTTCGTTCGGT
ATTTTTAATGGCGATGTTTTAGGGCTATCAGTTCGCGCATTAAAGACTAATAGCCATTCAAA
AATATTGTCTGTGCCACGTATTCTTACGCTTTCAGGTCAGAAGGGTTCTATCTCTGTTGGCC
AGAATGTCCCTTTTATTACTGGTCGTGTGACTGGTGAATCTGCCAATGTAAATAATCCATTT
CAGACGATTGAGCGTCAAAATGTAGGTATTTCCATGAGCGTTTTTCCTGTTGCAATGGCTGG

CGGTAATATTGTTCTGGATATTACCAGCAAGGCCGATAGTTTGAGTTCTTCTACTCAGGCAA
GTGATGTTATTACTAATCAAAGAAGTATTGCTACAACGGTTAATTTGCGTGATGGACAGACT
CTTTTACTCGGTGGCCTCACTGATTATAAAAACACTTCTCAAGATTCTGGCGTACCGTTCCT
GTCTAAAATCCCTTTAATCGGCCTCCTGTTTAGCTCCCGCTCTGATTCCAACGAGGAAAGCA
CGTTATACGTGCTCGTCAAAGCAACCATAGTACGCGCCCTGTAGCGGCGCATTAAGCGCGGC
GGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTT
TCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGG
GGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTT
GGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGG
AGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCG
GGCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCT
GATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTAAATATTTG
CTTATACAATCTTCCTGTTTTTGGGGCTTTTCTGATTATCAACCGGGGTACATATGATTGAC
ATGCTAGTTTTACGATTACCGTTCATCGATTCTCTTGTTTGCTCCAGACTCTCAGGCAATGA
CCTGATAGCCTTTGTAGACCTCTCAAAAATAGCTACCCTCTCCGGCATGAATTTATCAGCTA
GAACGGTTGAATATCATATTGATGGTGATTTGACTGTCTCCGGCCTTTCTCACCCTTTTGAA
TCTTTACCTACACATTACTCAGGCATTGCATTTAAAATATATGAGGGTTCTAAAAATTTTTA
TCCTTGCGTTGAAATAAAGGCTTCTCCCGCAAAAGTATTACAGGGTCATAATGTTTTTGGTA
CAACCGATTTAGCTTTATGCTCTGAGGCTTTATTGCTTAATTTTGCTAATTCTTTGCCTTGC
CTGTATGATTTATTGGATGTT (SEQ ID NO: 1)

GENE II: join(6006..6407,1..831)
translation=MIDMLVLRLPFIDSLVCSRLSGNDLIAFVDLSKIATLSGMNLSARTVEYHIDGD
LTVSGLSHPFESLPTHYSGIAFKIYEGSKNFYPCVEIKASPAKVLQGHNVFGTTDLALCSEA
LLLNFANSLPCLYDLLDVNATTISRIDATFSARAPNENIAKQVIDHLRNVSNGQTKSTRSQN
WESTVTWNETSRHRTLVAYLKHVELQHQIQQLSSKPSAKMTSYQKEQLKVLSNPDLLEFASG
LVRFEARIKTRYLKSFGLPLNLFDAIRFASDYNSQGKDLIFDLWSFSFSELFKAFEGDSMNI
YDDSAVLDAIQSKHFTITPSGKTSFAKASRYFGFYRRLVNEGYDSVALTMPRNSFWRYVSAL
VECGIPKSQLMNLSTCNNVVPLVRFINVDFSSQRPDWYNEPVLKIA (SEQ ID NO: 2)

GENE X: 496..831
translation=MNIYDDSAVLDAIQSKHFTITPSGKTSFAKASRYFGFYRRLVNEGYDSVALTMP
RNSFWRYVSALVECGIPKSQLMNLSTCNNVVPLVRFINVDFSSQRPDWYNEPVLKIA (SEQ
ID NO: 3)

GENE V:843..1106
translation=MIKVEIKPSQAQFTTRSGVSRQGKPYSLNEQLCYVDLGNEYPVLVKITLDEGQP
AYAPGLYTVHLSSFKVGQFGSLMIDRLRLVPAK (SEQ ID NO: 4)

GENE VII: 1108..1209
translation=MEQVADFDTIYQAMIQISVVLCFALGIIAGGQR (SEQ ID NO: 5)

GENE IX: 1206..1304
translation=MSVLVYSFASFVLGWCLRSGITYFTRLMETSS (SEQ ID NO: 6)

GENE VIII: 1301..1522
translation=MKKSLVLKASVAVATLVPMLSFAAEGDDPAKAAFNSLQASATEYIGYAWAMVVV
IVGATIGIKLFKKFTSKAS (SEQ ID NO: 7)

GENE III: 1579..2853
translation=MKKLLFAIPLVVPFYSHSAETVESCLAKPHTENSFTNVWKDDKTLDRYANYEGC
LWNATGVVVCTGDETQCYGTWVPIGLAIPENEGGGSEGGGSEGGGSEGGGTKPPEYGDTPIP
GYTYINPLDGTYPPGTEQNPANPNPSLEESQPLNTFMFQNNRFRNRQGALTVYTGTVTQGTD
PVKTYYQYTPVSSKAMYDAYWNGKFRDCAFHSGFNEDPFVCEYQGQSSDLPQPPVNAGGGSG
GGSGGGSEGGGSEGGGSEGGGSEGGGSGGGSGSGDFDYEKMANANKGAMTENADENALQSDA
KGKLDSVATDYGAAIDGFIGDVSGLANGNGATGDFAGSNSQMAQVGDGDNSPLMNNFRQYLP
SLPQSVECRPFVFSAGKPYEFSIDCDKINLFRGVFAFLLYVATFMYVFSTFANILRNKES
(SEQ ID NO: 8)

GENE VI: 2856..3194
translation=MPVLLGIPLLLRFLGFLLVTLFGYLLTFLKKGFGKIAIAISLFLALIIGLNSIL
VGYLSDISAQLPSDFVQGVQLILPSNALPCFYVILSVKAAIFIFDVKQKIVSYLDWDK (SEQ
ID NO: 9)

GENE I: 3196..4242
translation=MAVYFVTGKLGSGKTLVSVGKIQDKIVAGCKIATNLDLRLQNLPQVGRFAKTPR
VLRIPDKPSISDLLAIGRGNDSYDENKNGLLVLDECGTWFNTRSWNDKERQPIIDWFLHARK
LGWDIIFLVQDLSIVDKQARSALAEHVVYCRRLDRITLPFVGTLYSLITGSKMPLPKLHVGV
VKYGDSQLSPTVERWLYTGKNLYNAYDTKQAFSSNYDSGVYSYLTPYLSHGRYFKPLNLGQK
MKLTKIYLKKFSRVLCLAIGFASAFTYSYITQPKPEVKKVVSQTYDFDKFTIDSSQRLNLSY
RYVFKDSKGKLINSDDLQKQGYSLTYIDLCTVSIKKGNSNEIVKCN (SEQ ID NO: 10)

GENE IV: 4220..5500
translation=MKLLNVINFVFLMFVSSSSFAQVIEMNNSPLRDFVTWYSKQSGESVIVSPDVKG
TVTVYSSDVKPENLRNFFISVLRANNFDMVGSIPSIIQKYNPNNQDYIDELPSSDNQEYDDN
SAPSGGFFVPQNDNVTQTFKINNVRAKDLIRVVELFVKSNTSKSSNVLSIDGSNLLVVSAPK
DILDNLPQFLSTVDLPTDQILIEGLIFEVQQGDALDFSFAAGSQRGTVAGGVNTDRLTSVLS
SAGGSFGIFNGDVLGLSVRALKTNSHSKILSVPRILTLSGQKGSISVGQNVPFITGRVTGES
ANVNNPFQTIERQNVGISMSVFPVAMAGGNIVLDITSKADSLSSSTQASDVITNQRSIATTV
NLRDGQTLLLGGLTDYKNTSQDSGVPFLSKIPLIGLLFSSRSDSNEESTLYVLVKATIVRAL
(SEQ ID NO: 11)

[0090] Some embodiments of the protease PACE technology described herein utilize a "selection phage," a modified phage that comprises a gene of interest to be evolved and lacks a full-length gene encoding a protein required for the generation of infectious phage particles. In some such embodiments, the selection phage serves as the vector that replicates and evolves in the flow of host cells. For example, some M13 selection phages provided herein comprise a nucleic acid sequence encoding a protease to be evolved, *e.g.*, under the control of an M13 promoter, and lack all or part of a phage gene encoding a protein required for the generation of infectious phage particles, *e.g.*, gI, gII, gIII, gIV, gV, gVI, gVII, gVIII, gIX, or gX, or any combination thereof. For example, some M13 selection phages provided herein comprise a nucleic acid sequence encoding a protease to be evolved, *e.g.*, under the control of an M13 promoter, and lack all or part of a gene encoding a protein required for the

generation of infective phage particles, *e.g.*, the gIII gene encoding the pIII protein. An exemplary, non-limiting selection plasmid sequence, SP-MCS, comprising a multiple cloning site, into which a nucleic acid sequence encoding a protease to be evolved can be cloned, is provided below:

ATGATTGACATGCTAGTTTTACGATTACCGTTCATCGATTCTCTTGTTTGCTCCAGACTCTC
AGGCAATGACCTGATAGCCTTTGTAGACCTCTCAAAAATAGCTACCCTCTCCGGCATGAATT
TATCAGCTAGAACGGTTGAATATCATGTTGATGGTGATTTGACTGTCTCCGGCCTTTCTCAC
CCTTTTGAATCTTTACCTACACATTACTCAGGCATTGCATTTAAAATATATGAGGGTTCTAA
AAATTTTTATCCTTGCGTTGAAATAAAGGCTTCTCCCGCAAAAGTATTACAGGGTCATAATG
TTTTTGGTACAACCGATTTAGCTTTATGCTCTGAGGCTTTATTGCTTAATTTTGCTAATTCT
TTGCCTTGCCTGTATGATTTATTGGATGTTAACGCTACTACTATTAGTAGAATTGATGCCAC
CTTTTCAGCTCGCGCCCCAAATGAAAATATAGCTAAACAGGTTATTGACCATTTGCGAAATG
TATCTAATGGTCAAACTAAATCTACTCGTTCGCAGAATTGGGAATCAACTGTTACATGGAAT
GAAACTTCCAGACACCGTACTTTAGTTGCATATTTAAAACATGTTGAGCTACAGCACCAGAT
TCAGCAATTAAGCTCTAAGCCATCCGCAAAAATGACCTCTTATCAAAAGGAGCAATTAAAGG
TACTCTCTAATCCTGACCTGTTGGAGTTTGCTTCCGGGCTGGTTCGCTTTGAAGCTCGAATT
AGAACGCGATATTTGAAGTCTTTCGGGCTTCCTCTTAATCTTTTTGATGCAATCCGCTTTGC
TTCTGACTATAATAGTCAGGGTAAAGACCTGATTTTTGATTTATGGTCATTCTCGTTTTCTG
AACTGTTTAAAGCATTTGAGGGGGATTCAATGAATATTTATGACGATTCCGCAGTATTGGAC
GCTATCCAGTCTAAACATTTTACTATTACCCCCTCTGGCAAAACTTCTTTTGCAAAAGCCTC
TCGCTATTTTGGTTTTTATCGTCGTCTGGTAAACGAGGGTTATGATAGTGTTGCTCTTACTA
TGCCTCGTAATTCCTTTTGGCGTTATGTATCTGCATTAGTTGAATGTGGTATTCCTAAATCT
CAACTGATGAATCTTTCTACCTGTAATAATGTTGTTCCGTTAGTTCGTTTTATTAACGTAGA
TTTTTCTTCCCAACGTCCTGACTGGTATAATGAGCCAGTTCTTAAAATCGCATAAGGTAATT
CACAATGATTAAAGTTGAAATTAAACCATCTCAAGCCCAATTTACTACTCGTTCTGGTGTTT
CTCGTCAGGGCAAGCCTTATTCACTGAATGAGCAGCTTTGTTACGTTGATTTGGGTAATGAA
TATCCGGTTCTTGTCAAGATTACTCTTGATGAAGGTCAGCCAGCCTATGCGCCTGGTCTGTA
CACCGTTCATCTGTCCTCTTTCAAAGTTGGTCAGTTCGGTTCCCTTATGATTGACCGTCTGC
GCCTCGTTCCGGCTAAGTAACATGGAGCAGGTCGCGGATTTCGACACAATTTATCAGGCGAT
GATACAAATCTCCGTTGTACTTTGTTTCGCGCTTGGTATAATCGCTGGGGGTCAAAGATGAG
TGTTTTAGTGTATTCTTTCGCCTCTTTCGTTTTAGGTTGGTGCCTTCGTAGTGGCATTACGT
ATTTTACCCGTTTAATGGAAACTTCCTCATGAAAAAGTCTTTAGTCCTCAAAGCCTCTGTAG
CCGTTGCTACCCTCGTTCCGATGCTGTCTTTCGCTGCTGAGGGTGACGATCCCGCAAAAGCG
GCCTTTAACTCCCTGCAAGCCTCAGCGACCGAATATATCGGTTATGCGTGGGCGATGGTTGT
TGTCATTGTCGGCGCAACTATCGGTATCAAGCTGTTTAAGAAATTCACCTCGAAAGCAAGCT
GATAAACCGATACAATTAAAGGCTCCTTTTGGAGCCTTTTTTTCGCGCCAGAAGGAGACCA
AGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTCTGGAGA
TTTTCAACATGCTCCCTCAATCGGTTGAATGTCGCCCTTTTGTCTTTAGCGCTGGTAAACCA
TATGAATTTTCTATTGATTGTGACAAAATGAACTTATTCCGTGGTGTCTTTGCGTTTCTTTT
ATATGTTGCCACCTTTATGTATGTATTTTCTACGTTTGCTAACATACTGCGTAATAAGGAGT
CTTAATCATGCCAGTTCTTTTGGGTATTCCGTTATTATTGCGTTTCCTCGGTTTCCTTCTGG
TAACTTTGTTCGGCTATCTGCTTACTTTTCTTAAAAAGGGCTTCGGTAAGATAGCTATTGCT
ATTTCATTGTTTCTTGCTCTTATTATTGGGCTTAACTCAATTCTTGTGGGTTATCTCTCTGA
TATTAGCGCTCAATTACCCTCTGACTTTGTTCAGGGTGTTCAGTTAATTCTCCCGTCTAATG
CGCTTCCCTGTTTTTATGTTATTCTCTCTGTAAAGGCTGCTATTTTCATTTTTGACGTTAAA
CAAAAAATCGTTTCTTATTTGGATTGGGATAAATAATATGGCTGTTTATTTTGTAACTGGCA
AATTAGGCTCTGGAAAGACGCTCGTTAGCGTTGGTAAGATTCAGGATAAAATTGTAGCTGGG
TGCAAAATAGCAACTAATCTTGATTTAAGGCTTCAAAACCTCCCGCAAGTCGGGAGGTTCGC

TAAAACGCCTCGCGTTCTTAGAATACCGGATAAGCCTTCTATATCTGATTTGCTTGCTATTG
GGCGCGGTAATGATTCCTACGATGAAAATAAAAACGGCTTGCTTGTTCTCGATGAGTGCGGT
ACTTGGTTTAATACCCGTTCTTGGAATGATAAGGAAAGACAGCCGATTATTGATTGGTTTCT
ACATGCTCGTAAATTAGGATGGGATATTATTTTTCTTGTTCAGGACTTATCTATTGTTGATA
AACAGGCGCGTTCTGCATTAGCTGAACATGTTGTTTATTGTCGTCGTCTGGACAGAATTACT
TTACCTTTTGTCGGTACTTTATATTCTCTTATTACTGGCTCGAAAATGCCTCTGCCTAAATT
ACATGTTGGCGTTGTTAAATATGGCGATTCTCAATTAAGCCCTACTGTTGAGCGTTGGCTTT
ATACTGGTAAGAATTTGTATAACGCATATGATACTAAACAGGCTTTTTCTAGTAATTATGAT
TCCGGTGTTTATTCTTATTTAACGCCTTATTTATCACACGGTCGGTATTTCAAACCATTAAA
TTTAGGTCAGAAGATGAAATTAACTAAAATATATTTGAAAAAGTTTTCTCGCGTTCTTTGTC
TTGCGATTGGATTTGCATCAGCATTTACATATAGTTATATAACCCAACCTAAGCCGGAGGTT
AAAAAGGTAGTCTCTCAGACCTATGATTTTGATAAATTCACTATTGACTCTTCTCAGCGTCT
TAATCTAAGCTATCGCTATGTTTTCAAGGATTCTAAGGGAAAATTAATTAATAGCGACGATT
TACAGAAGCAAGGTTATTCACTCACATATATTGATTTATGTACTGTTTCCATTAAAAAAGGT
AATTCAAATGAAATTGTTAAATGTAATTAATTTTGTTTTCTTGATGTTTGTTTCATCATCTT
CTTTTGCTCAGGTAATTGAAATGAATAATTCGCCTCTGCGCGATTTTGTAACTTGGTATTCA
AAGCAATCAGGCGAATCCGTTATTGTTTCTCCCGATGTAAAAGGTACTGTTACTGTATATTC
ATCTGACGTTAAACCTGAAAATCTACGCAATTTCTTTATTTCTGTTTTACGTGCAAGTAATT
TTGATATGGTTGGTTCTAACCCTTCCATTATTCAGAAGTATAATCCAAACAATCAGGATTAT
ATTGATGAATTGCCATCATCTGATAATCAGGAATATGATGATAATTCCGCTCCTTCTGGTGG
TTTCTTTGTTCCGCAAAATGATAATGTTACTCAAACTTTTAAAATTAATAACGTTCGGGCAA
AGGATTTAATACGAGTTGTCGAATTGTTTGTAAAGTCTAATACTTCTAAATCCTCAAATGTA
TTATCTATTGACGGCTCTAATCTATTAGTTGTTAGTGCACCTAAAGATATTTTAGATAACCT
TCCTCAATTCCTTTCTACTGTTGATTTGCCAACTGACCAGATATTGATTGAGGGTTTGATAT
TTGAGGTTCAGCAAGGTGATGCTTTAGATTTTTCATTTGCTGCTGGCTCTCAGCGTGGCACT
GTTGCAGGCGGTGTTAATACTGACCGCCTCACCTCTGTTTTATCTTCTGCTGGTGGTTCGTT
CGGTATTTTTAATGGCGATGTTTTAGGGCTATCAGTTCGCGCATTAAAGACTAATAGCCATT
CAAAAATATTGTCTGTGCCACGTATTCTTACGCTTTCAGGTCAGAAGGGTTCTATCTTTGTT
GGCCAGAATGTCCCTTTTATTACTGGTCGTGTGACTGGTGAATCTGCCAATGTAAATAATCC
ATTTCAGACGATTGAGCGTCAAAATGTAGGTATTTCCATGAGCGTTTTTCCTGTTGCAATGG
CTGGCGGTAATATTGTTCTGGATATTACCAGCAAGGCCGATAGTTTGAGTTCTTCTACTCAG
GCAAGTGATGTTATTACTAATCAAAGAAGTACTGCTACAACGGTTAATTTGCGTGATGGACA
GACTCTTTTACTCGGTGGCCTCACTGATTATAAAAACACTTCTCAGGATTCTGGCGTACCGT
TCCTGTCTAAAATCCCTTTAATCGGCCTCCTGTTTAGCTCCCGCTCTGATTCTAACGAGGAA
AGCACGTTATACGTGCTCGTCAAAGCAACCATAGTACGCGCCCTGTAGCGGCGCATTAAGCG
CGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCT
CCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAA
TCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTG
ATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACG
TTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTAT
CTCGGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATG
AGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTAAATA
TTTGCTTATACAATCTTCCTGTTTTTGGGGCTTTTCTTATTATCAACCGGGGTACAT
(SEQ ID NO: 30)

[0091] One prerequisite for evolving proteases with a desired activity is to provide a selection system that confers a selective advantage to mutated protease variants exhibiting such an activity. The expression systems and fusion proteins comprising transcriptional

activators in an inactive form that are activated by protease activity thus constitute an important feature of some embodiments of the protease PACE technology provided herein.

[0092] In some embodiments, the host cell expresses a fusion molecule comprising (i) a transcriptional activator; and (ii) an inhibitor of the transcriptional activator of (i), wherein the inhibitor is fused to the transcriptional activator of (i) via a linker comprising a protease cleavage site that is cleaved by the protease of (a). In some embodiments, the host cell expresses a transcriptional activator that is fused to a polypeptide tag targeting the transcriptional activator for degradation, or directing export of the transcriptional activator from the host cell or into a compartment of the host cell in which the transcriptional activator cannot mediate transcription from a target promoter. Suitable transcriptional activators, inhibitors, and polypeptide tags will be readily apparent to those of skill in the art based on the instant disclosure.

[0093] In some embodiments, the transcriptional activator directly drives transcription from a target promoter. For example, in some such embodiments, the transcriptional activator may be an RNA polymerase. Suitable RNA polymerases and promoter sequences targeted by such RNA polymerases are well known to those of skill in the art. Exemplary suitable RNA polymerases include, but are not limited to, T7 polymerases (targeting T7 promoter sequences) and T3 RNA polymerases (targeting T3 promoter sequences). Additional suitable RNA polymerases will be apparent to those of skill in the art based on the instant disclosure, which is not limited in this respect.

[0094] In some embodiments, the transcriptional activator does not directly drive transcription, but recruits the transcription machinery of the host cell to a specific target promoter. Suitable transcriptional activators, such as, for example, Gal4 or fusions of the transactivation domain of the VP16 transactivator with DNA-binding domains, will be apparent to those of skill in the art based on the instant disclosure, and the disclosure is not limited in this respect.

[0095] In some embodiments, it is advantageous to link protease activity to enhanced phage packaging via a transcriptional activator that is not endogenously expressed in the host cells in order to minimize leakiness of the expression of the gene required for the generation of infectious phage particles through the host cell basal transcription machinery. For example, in some embodiments, it is desirable to drive expression of the gene required for the generation of infectious phage particles from a promoter that is not or is only minimally active in host cells in the absence of an exogenous transcriptional activator, and to provide the exogenous transcriptional activator, such as, for example, T7 RNA polymerase, as part of

the expression system linking protease activity to phage packaging efficiency. In some embodiments, the at least one gene for the generation of infectious phage particles is expressed in the host cells under the control of a promoter activated by the transcriptional activator, for example, under the control of a T7 promoter if the transcriptional activator is T7 RNA polymerase, and under the control of a T3 promoter if the transcriptional activator is T3 polymerase, and so on.

[0096] In some embodiments, the transcriptional activator is fused to an inhibitor that either directly inhibits or otherwise hinders the transcriptional activity of the transcriptional activator, for example, by directly interfering with DNA binding or transcription, by targeting the transcriptional activator for degradation through the host cells protein degradation machinery, or by directing export from the host cell or localization of the transcriptional activator into a compartment of the host cell in which it cannot activate transcription from its target promoter. In some embodiments, the inhibitor is fused to the transcriptional activator's N-terminus. In other embodiments, it is fused to the activator's C-terminus.

[0097] Typically, the fusion proteins employed in the protease PACE methods provided herein comprise a linker connecting the transcriptional activator to the inhibitor and comprising a protease cleavage site. In some embodiments, the transcriptional activity of the fusion protein is inhibited as compared to the activity of the transcriptional activator alone. Such inhibition can be detected by suitable assays well known to those of skill in the art. For example, a reporter assay for assessing the transcriptional activity of a fusion protein and a transcriptional activator alone may comprise a reporter expression construct comprising a nucleic acid encoding a reporter gene under the control of a promoter targeted by the transcriptional activator. The reporter gene may be, for example, a fluorescent protein or a protein that catalyzes a bioluminescent reaction. The transcriptional activity of the respective fusion protein or the transcriptional activator alone can then be determined by fluorescent or bioluminescent readouts. Additional suitable assays for determining the transcriptional activity of the fusion proteins provided herein will be apparent to those of skill in the art and include, without limitation, those described in International PCT Application, PCT/US2011/066747, filed December 22, 2011, published as WO 2012/088381 on June 28, 2012; and U.S. Application, U.S.S.N. 13/922,812, filed June 20, 2013, the entire contents of each of which are incorporated herein by reference.

[0098] In some embodiments, the transcriptional activity of the fusion protein is less than 50%, less than 25%, less than 20%, less than 10%, less than 5%, less than 4%, less than 3%, less than 1%, less than 0.1%, less than 0.01%, or less than 0.001% the activity of the

transcriptional activator alone, for example, as measured via a suitable assay, such as one of the exemplary, non-limiting fluorescent or bioluminescent assays provided herein or in International PCT Application, PCT/US2011/066747, filed December 22, 2011, published as WO 2012/088381 on June 28, 2012; and U.S. Application, U.S.S.N. 13/922,812, filed June 20, 2013, the entire contents of each of which are incorporated herein by reference.

[0099] In some embodiments, cleavage of the protease cleavage site comprised in the linker connecting the transcriptional activator and the inhibitor results in release of the inhibitor from the activator and thus activation of the transcriptional activator. In some embodiments, the activity of the transcriptional activator is increased at least 2-fold, at least 5-fold, at least 10-fold, at least 20-fold, at least 25-fold, at least 50-fold, at least 75-fold, at least 100-fold, at least 200-foled, at least 250-fold, at least 500-fold, at least 750-fold, at least 1000-fold, at least 2000-fold, at least 2500-fold, at least 5000-fold, at least 7500-fold, at least 10000-fold, or more upon cleavage of the protease cleavage site of the linker.

[00100] In some embodiments, the protease to be evolved is a therapeutic protease. Therapeutic proteases are well known to those of skill in the art and include, without limitation, proteases that are useful for the treatment of cardiovascular disease, sepsis, digestive disorders, inflammation, cystic fibrosis, retinal disorders, psoriasis, and other diseases. Exemplary therapeutic proteases include, without limitation, those proteases described in Craik *et al.*, Proteases as therapeutics. *Biochem J.* 2011 Apr 1;435(1):1-16, the entire contents of which are incorporated herein by reference. Some exemplary therapeutic proteases are listed below:

Protease	Indication
Urokinase (u-PA)	Thrombus, catheter clearing
t-PA (alteplase, Activase®)	Myocardial infarction, stroke
Reteplase (Retevase)	Myocardial infarction
TNK-tPA (tenecteplase, Metalyse®)	Myocardial infarction
Factor IX (BeneFIX®)	Haemophilia B
FVIIa (NovoSeven®)	Haemophilia A and B
Thrombin (Recothrom®)	Bleeding
Activated protein C (drotrecogin alfa, Xigris®)	Sepsis, septic shock
Botulinum toxin A (Botox®)	Muscle spasms
Botulinum toxin B (Myobloc)	Cervical dystonia
Zenpep® (pancrelipase)	Pancreatic Insufficiency
Liprotamase (*Aspergillus melleus*)	Cystic fibrosis with exocrine pancreatic

	insufficiency
Microplasmin	Vitreomacular adhesion
Glutamine-specific cysteine protease (EP-B2) Proline-specific prolylendopeptidase (PEP), ALV003	Coeliac disease
Microplasmin	Acute peripheral arterial occlusion, deep vein thrombosis
Plasmin	Peripheral arterial disease, thrombosis
Recombinant human lysosomal protease	Medical and cosmetic dermatological applications
Kallikrein	Thrombosis, peripheral vascular disease, cerebrovascular ischaemia
Calpain 3	Replacement therapy, calpainopathy
Penzyme	Psoriasis, eczema and dermatitis

Additional suitable therapeutic proteases will be apparent to those of skill in the art based on the instant disclosure.

[00101] In some embodiments, a therapeutic protease is subjected to PACE to improve the efficiency with which the protease cleaves its target sequence. This may allow achieving the same therapeutic effect with a smaller dose of administered protease or achieving a greater therapeutic effect with the same dose as compared to using a non-evolved protease. In some embodiments, the therapeutic protease may be subjected to PACE to decrease the cleavage of off-target cleavage sites, *e.g.*, by employing negative PACE selection strategies described in more detail elsewhere herein or otherwise known in the art. Decreasing off-target protease cleavage may decrease the toxicity and/or side effects of a therapeutic protease.

[00102] In some embodiments, the replicated vector isolated after a protease PACE experiment encodes a mutated protease that cleaves the protease cleavage site with higher efficiency and/or higher specificity than the original version of the protease used at the outset of the PACE experiment. The mutated protease can then be analyzed, and the critical mutations resulting in the observed improvements can be ascertained. For example, a plurality of different evolved proteases isolated from a single PACE experiment or from different PACE experiments starting with the same protease can be analyzed. Shared mutations amongst the evolved proteases are likely to contribute to enhanced protease activity, while mutations that occur only in individual proteases are not likely to contribute to enhanced protease activity. The contribution of an observed mutation after protease PACE may also be followed up in more depth, for example, by introducing the mutation into the

original protease and determining the level of improvement conferred by the respective mutation in isolation or in combination with other observed mutations. Those of skill in the art will know suitable methods for generating recombinant proteins to perform such follow-up experiments after protease PACE.

[00103] In some embodiments, the protease to be evolved is a target of a protease inhibitor. In some such embodiments, the protease inhibitor is a therapeutic agent. Protease inhibitors are an important class of therapeutic agents and are used, for example, in the treatment of viral infections, protozoal infections, and certain types of cancer. Antiviral protease inhibitors are used, for example, in the context of retroviral infections, and some examples of known antiretroviral protease inhibitors include, without limitation, Saquinavir (*e.g.*, Fortovase, Invirase; see, *e.g.*, U.S. Patent 5,196,438); Ritonavir (*e.g.*, Norvir; see, *e.g.*, U.S. Patent 5,541,206); Indinavir (*e.g.*, Crixivan; see, *e.g.*, U.S. Patent 5,413,999); Nelfinavir(*e.g.*, Viracept; see, *e.g.*, U.S. Patent 5,484,926); Amprenavir (*e.g.*, Agenerase; see, *e.g.*, U.S. Patent 5,585,397); Lopinavir (*e.g.*, Kaletra; see, *e.g.*, U.S. Patent 5,914,332); Atazanavir (*e.g.*, Reyataz; see, *e.g.*, U.S. Patent 5,849,911); Fosamprenavir (*e.g.*, Lexiva, Telzir); Tipranavir (*e.g.*, Aptivus); Darunavir (*e.g.*, Prezista, Tibotec; U.S. Patent 6,248,775); Simeprevir (*e.g.*, Olysio, TMC435; U.S. Patent 7,671,032); Danoprevir; and Asunaprevir. Additional suitable protease inhibitors will be apparent to those of skill in the art based on the instant disclosure, which is not limited in this respect.

[00104] In some embodiments, protease PACE is carried out in the presence of the protease inhibitor during all or some of the steps in which the phage encoding the protease mutates and replicates in the flow of host cells. In some embodiments, the protease inhibitor is used at a concentration that does not completely inhibit the activity of the protease. This results, in some embodiments, in the selective pressure being relieved to a certain extent, which allows a diversification of the protease pool since even mutated variants of the protease that are not active in the presence of the inhibitor will be able to propagate in the flow of host cells at a basal level, *e.g.*, at a level that prevents dilution and/or washout of such protease variants from the population of vectors replicating in the cell flow. Such a diversification may be advantageous in some embodiments, *e.g.*, to provide a more varied initial pool of proteases that can then be subjected to evolution under more stringent selective pressure, *e.g.*, in the presence of a concentration of protease inhibitor that results in complete inactivation of the activity of the original protease. Relieving the selective pressure may also be performed at one or more time points during an ongoing PACE experiment, for example,

to achieve re-diversification of an evolved protease pool in order to access additional mutations that could not be reached under constant stringent selection pressure.

[00105] For example, in some embodiments, a PACE experiment may include an initial phase, *e.g.*, of 1 hour, 2 hours, 3 hours, 4 hours, 5 hours, or 6 hours, or longer, in which now or only low selective pressure is applied, for example, by providing host cells in which the selection phage can replicate regardless of any advantageous mutations, and a subsequent phase, *e.g.*, of 1 hour, 2 hours, 3 hours, 4 hours, 5 hours, or 6 hours, or longer, in which increased selection pressure is applied, for example, by providing host cells that preferentially support replication of phage comprising beneficial mutations, e.g., host cells harboring an accessory plasmid that links protease activity to phage replication efficiency, as described herein. In some PACE experiments, the selective pressure is increased over time, e.g., by subsequently using more and more selective accessory plasmids, e.g., starting with high copy number accessory plasmids and moving to plasmids with a lower copy number. In some embodiments, phases of no or low stringency of selection are alternated with phases of high selective pressure, in order to create multiple rounds of library diversification and selection.

[00106] In some embodiments, the protease inhibitor is used at a concentration that results in an inhibition of at least 50%, at least 75%, at least 80%, at least 90%, at least 95%, at least 97.5%, at least 98%, at least 99%, at least 99.5%, or at least 99.9% of the protease activity. In some embodiments, the protease inhibitor is used at an initial concentration that does not completely inhibit the activity of the protease, and the concentration of the inhibitor is subsequently raised to increase the level of protease inhibition in order to increase the stringency of selective pressure during the PACE experiment.

[00107] In some embodiments in which a protease inhibitor is used, a phage vector isolated at the end of a PACE experiment encodes a mutated protease variant that cleaves the protease cleavage site in the presence of the inhibitor. In other words, in some embodiments, a protease variant that is resistant to the protease inhibitor used is isolated. In some embodiments, the vector isolated at the end of a PACE experiment encodes a mutated protease that cleaves the protease cleavage site in the presence of the inhibitor with higher efficiency than the original version of the protease used at the outset of the PACE experiment.

[00108] In some embodiments, the protease evolution methods provided herein comprise an initial or intermittent phase of diversifying the population of vectors by mutagenesis, in which the cells are incubated under conditions suitable for mutagenesis of the gene encoding the protease in the absence of stringent selection or in the absence of any

selection for evolved protease variants that have acquired a desired activity. Such low-stringency selection or no selection periods may be achieved by supporting expression of the gene for the generation of infectious phage particles in the absence of desired protease activity, for example, by providing an inducible expression construct comprising a gene encoding the respective packaging protein under the control of an inducible promoter and incubating under conditions that induce expression of the promoter, *e.g.*, in the presence of the inducing agent. Suitable inducible promoters and inducible expression systems are described herein and in International PCT Application, PCT/US2011/066747, filed December 22, 2011, published as WO 2012/088381 on June 28, 2012; and U.S. Application, U.S.S.N. 13/922,812, filed June 20, 2013, the entire contents of each of which are incorporated herein by reference. Additional suitable promoters and inducible gene expression systems will be apparent to those of skill in the art based on the instant disclosure. In some embodiments, the method comprises a phase of stringent selection for a mutated protease version. If an inducible expression system is used to relieve selective pressure, the stringency of selection can be increased by removing the inducing agent from the population of cells in the lagoon, thus turning expression from the inducible promoter off, so that any expression of the gene required for the generation of infectious phage particles must come from the protease activity-dependent expression system.

[00109] One aspect of the PACE protease evolution methods provided herein is the mutation of the initially provided vectors encoding a protease of interest. In some embodiments, the host cells within the flow of cells in which the vector replicates are incubated under conditions that increase the natural mutation rate. This may be achieved by contacting the host cells with a mutagen, such as certain types of radiation or to a mutagenic compound, or by expressing genes known to increase the cellular mutation rate in the cells.

[00110] In some embodiments, the host cells are exposed to a mutagen. Typically, the concentration of the mutagen will be chosen to maximize the mutation rate while not being toxic to the host cells during the retention time in the lagoon. Ideally, a mutagen is used at a concentration or level of exposure that induces a desired mutation rate in a given host cell or viral vector population, but is not significantly toxic to the host cells used within the average time frame a host cell is exposed to the mutagen or the time a host cell is present in the host cell flow before being replaced by a fresh host cell. In some embodiments, the mutagen is ionizing radiation, ultraviolet radiation, base analogs, deaminating agents (*e.g.*, nitrous acid), intercalating agents (*e.g.*, ethidium bromide), alkylating agents (*e.g.*, ethylnitrosourea), transposons, bromine, azide salts, psoralen, benzene,3- Chloro-4-(dichloromethyl)-5-

hydroxy-2(5H)-furanone (MX) (CAS no. 77439-76-0), O,O-dimethyl-S-(phthalimidomethyl)phosphorodithioate (phos-met) (CAS no. 732-11- 6), formaldehyde (CAS no. 50-00-0), 2-(2-furyl)-3-(5-nitro-2-furyl)acrylamide (AF-2) (CAS no. 3688-53-7), glyoxal (CAS no. 107-22-2), 6-mercaptopurine (CAS no. 50-44- 2), N-(trichloromethylthio)-4-cyclohexane-l,2-dicarboximide (captan) (CAS no. 133- 06-2), 2-aminopurine (CAS no. 452-06-2), methyl methane sulfonate (MMS) (CAS No. 66-27-3), 4-nitroquinoline 1 -oxide (4-NQO) (CAS No. 56-57-5), N4-Aminocytidine (CAS no. 57294-74-3), sodium azide (CAS no. 26628-22-8), N-ethyl-N-nitrosourea (ENU) (CAS no. 759-73-9), N-methyl-N-nitrosourea (MNU) (CAS no. 820-60-0), 5- azacytidine (CAS no. 320-67-2), cumene hydroperoxide (CHP) (CAS no. 80-15-9), ethyl methanesulfonate (EMS) (CAS no. 62-50-0), N-ethyl-N -nitro-N-nitrosoguanidine (ENNG) (CAS no. 4245-77-6), N-methyl-N -nitro-N-nitrosoguanidine (MNNG) (CAS no. 70-25-7), 5-diazouracil (CAS no. 2435-76-9), or t-butyl hydroperoxide (BHP) (CAS no. 75-91-2). Additional suitable mutagens will be known to those of skill in the art, and the disclosure is not limited in this respect.

[00111] In some embodiments, the mutation rate of the host cells is increased by contacting the cells with a mutagenesis plasmid encoding gene products known to increase the frequency of mutations in the host cells. In some embodiments, the host cells are contacted with a mutagenesis plasmid. In some embodiments, the mutagenesis plasmid comprises a gene expression cassette encoding a mutagenesis-promoting gene product. In some embodiments, the mutagenesis plasmid comprises a gene expression cassette encoding *umuC* (a component of *E. coli* translesion synthesis polymerase V, e.g., as set forth in GenBank M10107.1), *dam* (deoxyadenosine methylase, e.g., as set forth in GenBank J01600.1), or *seqA* (a hemimethylated-GATC binding domain, e.g., as set forth in GenBank U07651.1), or any combination thereof. In some embodiments, the mutagenesis plasmid further comprises a nucleic acid encoding *UmuD'*, and/or *RecA*.

[00112] In some embodiments, the mutagenesis-promoting gene is under the control of an inducible promoter. Suitable inducible promoters are well known to those of skill in the art and include, for example, arabinose-inducible promoters, tetracycline or doxycyclin-inducible promoters, and tamoxifen-inducible promoters. In some embodiments, the host cell population is contacted with an inducer of the inducible promoter in an amount sufficient to effect an increased rate of mutagenesis. For example, in some embodiments, a bacterial host cell population is provided in which the host cells comprise a mutagenesis plasmid in which a *umuC*, *dam*, and *seqA* expression cassette is controlled by an arabinose-inducible promoter.

In some such embodiments, the population of host cells is contacted with the inducer, for example, arabinose, in an amount sufficient to induce an increased rate of mutation.

[00113] The use of an inducible mutagenesis plasmid allows one to generate a population of fresh, uninfected host cells in the absence of the inducer, thus avoiding an increased rate of mutation in the fresh host cells before they are introduced into the population of cells contacted with the viral vector. Once introduced into this population, however, these cells can then be induced to support an increased rate of mutation, which is particularly useful in some embodiments of continuous evolution. For example, in some embodiments, the host cell comprises a mutagenesis plasmid as described herein, which includes an arabinose-inducible promoter driving expression of *umuC*, *dam*, and *seqA* from a pBAD promoter (see, *e.g.*, Khlebnikov A, Skaug T, Keasling JD. *Modulation of gene expression from the arabinose-inducible araBAD promoter.* J Ind Microbiol Biotechnol. 2002 Jul;29(1):34-7; incorporated herein by reference for disclosure of a pBAD promoter). In some embodiments, the fresh host cells are not exposed to arabinose, which activates expression of the above-identified genes and, thus, increases the rate of mutations in the arabinose-exposed cells, until the host cells reach the lagoon in which the population of selection phage replicates. Accordingly, in some embodiments, the mutation rate in the host cells is normal until they become part of the host cell population in the lagoon, where they are exposed to the inducer (*e.g.*, arabinose) and, thus , to increased mutagenesis. In some embodiments, a method of continuous evolution is provided that includes a phase of diversifying the population of viral vectors by mutagenesis, in which the cells are incubated under conditions suitable for mutagenesis of the viral vector in the absence of stringent selection for the mutated replication product of the viral vector encoding the evolved protein. This is particularly useful in embodiments in which a desired function to be evolved is not merely an increase in an already present function, for example, an increase in the cleavage activity of a protease towards its original cleavage site, but the acquisition of a function not present in the protease to be evolved at the outset of the evolution procedure, such as, for example, cleavage of a target site not recognized by the original version of the protease to be evolved. A step of diversifying the pool of mutated versions of the gene of interest within the population of viral vectors, for example, of phage, allows for an increase in the chance to find a mutation that conveys the desired function, e.g., new substrate specificity in a protease to be evolved.

[00114] In some embodiments, the host cell population is contacted with a mutagen continuously during a PACE experiment. In other embodiments, the host cell population is

contacted with the mutagen intermittently, creating phases of increased mutagenesis, and accordingly, of increased viral vector diversification. For example, in some embodiments, the host cells are exposed to a concentration of mutagen sufficient to generate an increased rate of mutagenesis in the gene of interest for about 10%, about 20%, about 30%, about 40%, about 50%, or about 75% of the time. In embodiments where a mutagenesis plasmid is employed, intermittent exposure to the encoded mutagenesis-increasing gene products can be achieved by using inducible promoters and adding or withdrawing the inducing agent from the host cell culture media during the PACE experiment.

[00115] In some embodiments, the link between desired protease activity and selective advantage for an encoding phage is provided by an expression system in which at least one gene for the generation of infectious phage particles is expressed in response to the desired protease activity as described in more detail elsewhere herein. In some embodiments, the at least one gene for the generation of infectious phage particles to another host cell is a gene required for the production of infectious phage particles. In some embodiments, the vector is M13 phage, and the at least one gene for the generation of infectious phage particles comprises a full-length M13 pIII gene. In some embodiments, the host cells comprise an accessory plasmid comprising an expression construct encoding the at least one gene for the generation of infectious phage particles, *e.g.*, the full-length pIII protein, under the control of a conditional promoter that is activated by the transcriptional activator comprised in the fusion protein.

[00116] In some embodiments, the conditional promoter of the accessory plasmid is a promoter, the transcriptional activity of which can be regulated over a wide range, for example, over 2, 3, 4, 5, 6, 7, 8, 9, or 10 orders of magnitude by the activating function, for example, the function of a protease of interest. In some embodiments, the conditional promoter has a basal activity that allows for baseline packaging of viral vectors even in the absence of the desired protease activity or in the presence of only minimal desired protease activity. This allows for starting a continuous evolution process with a viral vector population comprising versions of the protease of interest that only show minimal activation of the conditional promoter, *e.g.*, of wild-type proteases that are fully inhibited by a therapeutic protease inhibitor. In the process of continuous evolution, any mutation in the gene of interest that increases activity of the conditional promoter directly translates into higher expression levels of the gene required for the generation of infectious viral particles in the host cells harboring the vector comprising such a mutation, and, thus, into a competitive

advantage over other viral vectors carrying minimally active or loss-of-function versions of the protease of interest.

[00117] One function of the accessory plasmid is to provide a gene for the generation of infectious phage particles under the control of a conditional promoter the activity of which depends on a function of the protease of interest. Accordingly, the accessory plasmid provides a selection mechanism that favors desirable mutations over inconsequential mutations or mutations that are detrimental to the desired function. The stringency of selective pressure imposed by the accessory plasmid in a continuous evolution procedure as provided herein can be modulated. For example, an accessory plasmid may be used at different copy numbers per cell, may comprise a conditional promoter having a base line transcription rate ("leakiness") that prevents washout of unmutated sequences from the lagoon while still providing a selective advantage to desirable mutations. In some embodiments, an accessory plasmid comprising an expression cassette in which the gene for the generation of infectious phage particles is under the control of an inducible promoter that can be activated by a chemical compound (*e.g.*, a tet-on promoter), allowing for titration of the expression of the gene for the generation of infectious phage particles during a continuous evolution experiment.

[00118] In some embodiments, the use of low copy number accessory plasmids results in an elevated stringency of selection for versions of the gene encoding a protease of interest that activate the conditional promoter on the accessory plasmid, while the use of high copy number accessory plasmids results in a lower stringency of selection. The copy number of an accessory plasmid will depend to a large part on the origin of replication employed. Those of skill in the art will be able to determine suitable origins of replication in order to achieve a desired copy number. The following table lists some non-limiting examples of vectors of different copy numbers and with different origins of replication:

Plasmids	Origin of Replication	Copy number	Class
pUC vectors	pMB1*	500-700	high copy
pBluescript® vectors	ColE1	300-500	high copy
pGEM® vectors	pMB1*	300-400	high copy
pTZ vectors	pMB1*	>1000	high copy
pBR322 and derivatives	pMB1*	15-20	low copy
pACYC and derivatives	p15A	10-12	low copy

pSC101 and derivatives	pSC101	~5	very low copy

*The pMB1 origin of replication is closely related to that of ColE1 and falls in the same incompatibility group. The high-copy plasmids listed here contain mutated versions of this origin.

[00119] It should be understood that one function of the accessory plasmid, namely to provide a gene for the generation of infectious phage particles under the control of a conditional promoter, the activity of which depends on a function of the gene of interest, can be conferred to a host cell in alternative ways. Such alternatives include, but are not limited to, permanent insertion of a gene construct comprising the conditional promoter and the respective gene into the genome of the host cell, or introducing it into the host cell using a different vector, for example, a phagemid, a cosmid, a phage, a virus, or an artificial chromosome. Additional ways to confer accessory plasmid function to host cells will be evident to those of skill in the art, and the invention is not limited in this respect.

[00120] The sequences of two exemplary, non-limiting accessory plasmids, AP-MCS-A, and AP-MCS-P, respectively, are provided below:

AP-MCS-A:

GGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAG
CGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGC
CGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTA
GGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTA
TCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAAC
GTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCA
AATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTACTCTGCT
AGCAAGTAAGGCCGACAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGTACCG
AGCTCGAATTCCCTTTTTTTTTGGAGATTTTCAACGTGAAAAAATTATTATTCGCAATTCCT
TTAGTTGTTCCTTTCTATTCTCACTCCGCTGAAACTGTTGAAAGTTGTTTAGCAAAACCCCA
TACAGAAAATTCATTTACTAACGTCTGGAAAGACGACAAAACTTTAGATCGTTACGCTAACT
ATGAGGGCTGTCTGTGGAATGCTACAGGCGTTGTAGTTTGTACTGGTGACGAAACTCAGTGT
TACGGTACATGGGTTCCTATTGGGCTTGCTATCCCTGAAAATGAGGGTGGTGGCTCTGAGGG
TGGCGGTTCTGAGGGTGGCGGTTCTGAGGGTGGCGGTACTAAACCTCCTGAGTACGGTGATA
CACCTATTCCGGGCTATACTTATATCAACCCTCTCGACGGCACTTATCCGCCTGGTACTGAG
CAAAACCCCGCTAATCCTAATCCTTCTCTTGAGGAGTCTCAGCCTCTTAATACTTTCATGTT
TCAGAATAATAGGTTCCGAAATAGGCAGGGGGCATTAACTGTTTATACGGGCACTGTTACTC
AAGGCACTGACCCCGTTAAAACTTATTACCAGTACACTCCTGTATCATCAAAAGCCATGTAT
GACGCTTACTGGAACGGTAAATTCAGAGACTGCGCTTTCCATTCTGGCTTTAATGAGGATCC
ATTCGTTTGTGAATATCAAGGCCAATCGTCTGACCTGCCTCAACCTCCTGTCAATGCTGGCG
GCGGCTCTGGTGGTGGTTCTGGTGGCGGCTCTGAGGGTGGTGGCTCTGAGGGTGGCGGTTCT
GAGGGTGGCGGCTCTGAGGGAGGCGGTTCCGGTGGTGGCTCTGGTTCCGGTGATTTTGATTA
TGAAAAGATGGCAAACGCTAATAAGGGGGCTATGACCGAAAATGCCGATGAAAACGCGCTAC
AGTCTGACGCTAAAGGCAAACTTGATTCTGTCGCTACTGATTACGGTGCTGCTATCGATGGT
TTCATTGGTGACGTTTCCGGCCTTGCTAATGGTAATGGTGCTACTGGTGATTTTGCTGGCTC
TAATTCCCAAATGGCTCAAGTCGGTGACGGTGATAATTCACCTTTAATGAATAATTTCCGTC

AATATTTACCTTCCCTCCCTCAATCGGTTGAATGTCGCCCTTTTGTCTTTGGCGCTGGTAAA
CCATATGAATTTTCTATTGATTGTGACAAAATAAACTTATTCCGTGGTGTCTTTGCGTTTCT
TTTATATGTTGCCACCTTTATGTATGTATTTTCTACGTTTGCTAACATACTGCGTAATAAGG
AGTCTTAATCATGCCAGTTCTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTT
TTTTGCCTTGTCGGCCTTACTTGCTAAATACATTCAAATATGTATCCGCTCATGAGACAATA
ACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTG
TCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTG
GTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCT
CAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTT
TTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGT
CGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCT
TACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTG
CGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAAC
ATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAA
CGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTG
GCGAACTACTTACTCTAGCTTCCCGGCAACAATTGATAGACTGGATGGAGGCGGATAAAGTT
GCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGC
CGGTGAGCGTGGCTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTA
TCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
GAGATAGGTGCCTCACTGATTAAGCATTGGTAAGAACCTCAGATCCTTCCGTGATGGTAACT
TCACTAGTTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTT
AACGTGAGTTTTCGTTCCACTGAGCGTCAGAGAACCTCAGATCCTTCCGTATTTAGCCAGTA
TGTTCTCTAGTGTGGTTCGTTGTTTTTGCGTGAGCCATGAGAACGAACCATTGAGATCATGC
TTACTTTGCATGTCACTCAAAAATTTTGCCTCAAAACTGGTGAGCTGAATTTTTGCAGTTAA
AGCATCGTGTAGTGTTTTTCTTAGTCCGTTACGTAGGTAGGAATCTGATGTAATGGTTGTTG
GTATTTTGTCACCATTCATTTTTATCTGGTTGTTCTCAAGTTCGGTTACGAGATCCATTTGT
CTATCTAGTTCAACTTGGAAAATCAACGTATCAGTCGGGCGGCCTCGCTTATCAACCACCAA
TTTCATATTGCTGTAAGTGTTTAAATCTTTACTTATTGGTTTCAAAACCCATTGGTTAAGCC
TTTTAAACTCATGGTAGTTATTTTCAAGCATTAACATGAACTTAAATTCATCAAGGCTAATC
TCTATATTTGCCTTGTGAGTTTTCTTTTGTGTTAGTTCTTTTAATAACCACTCATAAATCCT
CATAGAGTATTTGTTTTCAAAAGACTTAACATGTTCCAGATTATATTTTATGAATTTTTTTA
ACTGGAAAAGATAAGGCAATATCTCTTCACTAAAAACTAATTCTAATTTTTCGCTTGAGAAC
TTGGCATAGTTTGTCCACTGGAAAATCTCAAAGCCTTTAACCAAAGGATTCCTGATTTCCAC
AGTTCTCGTCATCAGCTCTCTGGTTGCTTTAGCTAATACACCATAAGCATTTTCCCTACTGA
TGTTCATCATCTGAGCGTATTGGTTATAAGTGAACGATACCGTCCGTTCTTTCCTTGTAGGG
TTTTCAATCGTGGGGTTGAGTAGTGCCACACAGCATAAAATTAGCTTGGTTTCATGCTCCGT
TAAGTCATAGCGACTAATCGCTAGTTCATTTGCTTTGAAAACAACTAATTCAGACATACATC
TCAATTGGTCTAGGTGATTTTAATCACTATACCAATTGAGATGGGCTAGTCAATGATAATTA
CTAGTCCTTTTCCTTTGAGTTGTGGGTATCTGTAAATTCTGCTAGACCTTTGCTGGAAAACT
TGTAAATTCTGCTAGACCCTCTGTAAATTCCGCTAGACCTTTGTGTGTTTTTTTTGTTTATA
TTCAAGTGGTTATAATTTATAGAATAAAGAAAGAATAAAAAAAGATAAAAAGAATAGATCCC
AGCCCTGTGTATAACTCACTACTTTAGTCAGTTCCGCAGTATTACAAAAGGATGTCGCAAAC
GCTGTTTGCTCCTCTACAAAACAGACCTTAAAACCCTAAAGGCTTAAGTAGCACCCTCGCAA
GCTCGGGCAAATCGCTGAATATTCCTTTTGTCTCCGACCATCAGGCACCTGAGTCGCTGTCT
TTTTCGTGACATTCAGTTCGCTGCGCTCACGGCTCTGGCAGTGAATGGGGGTAAATGGCACT
ACAGGCGCCTTTTATGGATTCATGCAAGGAAACTACCCATAATACAAGAAAAGCCCGTCACG
GGCTTCTCAGGGCGTTTTATGGCGGGTCTGCTATGTGGTGCTATCTGACTTTTTGCTGTTCA
GCAGTTCCTGCCCTCTGATTTTCCAGTCTGACCACTTCGGATTATCCCGTGACAGGTCATTC
AGACTGGCTAATGCACCCAGTAAGGCAGCGGTATCATCAACT (SEQ ID NO: 31)

AP-MCS-P:

GGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAG
CGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGC
CGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTA
GGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTA
TCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAAC
GTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCA
AATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTACTCTGCT
AGCAAGTAAGGCCGACAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGTACCG
AGCTCGAATTCCCTTTTTTTTTGGAGATTTTCAACGTGAAAAAATTATTATTCGCAATTCCT
TTAGTTGTTCCTTTCTATTCTCACTCCGCTGAAACTGTTGAAAGTTGTTTAGCAAAACCCCA
TACAGAAAATTCATTTACTAACGTCTGGAAAGACGACAAAACTTTAGATCGTTACGCTAACT
ATGAGGGCTGTCTGTGGAATGCTACAGGCGTTGTAGTTTGTACTGGTGACGAAACTCAGTGT
TACGGTACATGGGTTCCTATTGGGCTTGCTATCCCTGAAAATGAGGGTGGTGGCTCTGAGGG
TGGCGGTTCTGAGGGTGGCGGTTCTGAGGGTGGCGGTACTAAACCTCCTGAGTACGGTGATA
CACCTATTCCGGGCTATACTTATATCAACCCTCTCGACGGCACTTATCCGCCTGGTACTGAG
CAAAACCCCGCTAATCCTAATCCTTCTCTTGAGGAGTCTCAGCCTCTTAATACTTTCATGTT
TCAGAATAATAGGTTCCGAAATAGGCAGGGGGCATTAACTGTTTATACGGGCACTGTTACTC
AAGGCACTGACCCCGTTAAAACTTATTACCAGTACACTCCTGTATCATCAAAAGCCATGTAT
GACGCTTACTGGAACGGTAAATTCAGAGACTGCGCTTTCCATTCTGGCTTTAATGAGGATCC
ATTCGTTTGTGAATATCAAGGCCAATCGTCTGACCTGCCTCAACCTCCTGTCAATGCTGGCG
GCGGCTCTGGTGGTGGTTCTGGTGGCGGCTCTGAGGGTGGTGGCTCTGAGGGTGGCGGTTCT
GAGGGTGGCGGCTCTGAGGGAGGCGGTTCCGGTGGTGGCTCTGGTTCCGGTGATTTTGATTA
TGAAAAGATGGCAAACGCTAATAAGGGGGCTATGACCGAAAATGCCGATGAAAACGCGCTAC
AGTCTGACGCTAAAGGCAAACTTGATTCTGTCGCTACTGATTACGGTGCTGCTATCGATGGT
TTCATTGGTGACGTTTCCGGCCTTGCTAATGGTAATGGTGCTACTGGTGATTTTGCTGGCTC
TAATTCCCAAATGGCTCAAGTCGGTGACGGTGATAATTCACCTTTAATGAATAATTTCCGTC
AATATTTACCTTCCCTCCCTCAATCGGTTGAATGTCGCCCTTTTGTCTTTGGCGCTGGTAAA
CCATATGAATTTTCTATTGATTGTGACAAAATAAACTTATTCCGTGGTGTCTTTGCGTTTCT
TTTATATGTTGCCACCTTTATGTATGTATTTTCTACGTTTGCTAACATACTGCGTAATAAGG
AGTCTTAATCATGCCAGTTCTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTT
TTTTGCCTTGTCGGCCTTACTTGCTAAATACATTCAAATATGTATCCGCTCATGAGACAATA
ACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTG
TCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTG
GTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCT
CAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTT
TTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGT
CGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCT
TACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTG
CGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAAC
ATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAA
CGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTG
GCGAACTACTTACTCTAGCTTCCCGGCAACAATTGATAGACTGGATGGAGGCGGATAAAGTT
GCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGC
CGGTGAGCGTGGCTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTA
TCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
GAGATAGGTGCCTCACTGATTAAGCATTGGTAAGAACCTCAGATCCTTCCGTGATGGTAACT
TCACTAGTTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTT
AACGTGAGTTTTCGTTCCACTGAGCGTCAGAGAACCTCAGATCCTTCCGTATTTAGCCAGTA
TGTTCTCTAGTGTGGTTCGTTGTTTTTGCGTGAGCCATGAGAACGAACCATTGAGATCATGC

TTACTTTGCATGTCACTCAAAAATTTTGCCTCAAAACTGGTGAGCTGAATTTTTGCAGTTAA
AGCATCGTGTAGTGTTTTTCTTAGTCCGTTACGTAGGTAGGAATCTGATGTAATGGTTGTTG
GTATTTTGTCACCATTCATTTTTATCTGGTTGTTCTCAAGTTCGGTTACGAGATCCATTTGT
CTATCTAGTTCAACTTGGAAAATCAACGTATCAGTCGGGCGGCCTCGCTTATCAACCACCAA
TTTCATATTGCTGTAAGTGTTTAAATCTTTACTTATTGGTTTCAAAACCCATTGGTTAAGCC
TTTTAAACTCATGGTAGTTATTTTCAAGCATTAACATGAACTTAAATTCATCAAGGCTAATC
TCTATATTTGCCTTGTGAGTTTTCTTTTGTGTTAGTTCTTTTAATAACCACTCATAAATCCT
CATAGAGTATTTGTTTTCAAAAGACTTAACATGTTCCAGATTATATTTTATGAATTTTTTTA
ACTGGAAAAGATAAGGCAATATCTCTTCACTAAAAACTAATTCTAATTTTTCGCTTGAGAAC
TTGGCATAGTTTGTCCACTGGAAAATCTCAAAGCCTTTAACCAAAGGATTCCTGATTTCCAC
AGTTCTCGTCATCAGCTCTCTGGTTGCTTTAGCTAATACACCATAAGCATTTTCCCTACTGA
TGTTCATCATCTGAGCGTATTGGTTATAAGTGAACGATACCGTCCGTTCTTTCCTTGTAGGG
TTTTCAATCGTGGGGTTGAGTAGTGCCACACAGCATAAAATTAGCTTGGTTTCATGCTCCGT
TAAGTCATAGCGACTAATCGCTAGTTCATTTGCTTTGAAAACAACTAATTCAGACATACATC
TCAATTGGTCTAGGTGATTTTAATCACTATACCAATTGAGATGGGCTAGTCAATGATAATTA
CTAGTCCTTTTCCTTTGAGTTGTGGGTATCTGTAAATTCTGCTAGACCTTTGCTGGAAAACT
TGTAAATTCTGCTAGACCCTCTGTAAATTCCGCTAGACCTTTGTGTGTTTTTTTTGTTTATA
TTCAAGTGGTTATAATTTATAGAATAAAGAAAGAATAAAAAAAGATAAAAAGAATAGATCCC
AGCCCTGTGTATAACTCACTACTTTAGTCAGTTCCGCAGTATTACAAAAGGATGTCGCAAAC
GCTGTTTGCTCCTCTACAAAACAGACCTTAAAACCCTAAAGGCTTAAGTAGCACCCTCGCAA
GCTCGGGCAAATCGCTGAATATTCCTTTTGTCTCCGACCATCAGGCACCTGAGTCGCTGTCT
TTTTCGTGACATTCAGTTCGCTGCGCTCACGGCTCTGGCAGTGAATGGGGGTAAATGGCACT
ACAGGCGCCTTTTATGGATTCATGCAAGGAAACTACCCATAATACAAGAAAAGCCCGTCACG
GGCTTCTCAGGGCGTTTTATGGCGGGTCTGCTATGTGGTGCTATCTGACTTTTTGCTGTTCA
GCAGTTCCTGCCCTCTGATTTTCCAGTCTGACCACTTCGGATTATCCCGTGACAGGTCATTC
AGACTGGCTAATGCACCCAGTAAGGCAGCGGTATCATCAACT (SEQ ID NO: 32)

[00121] Since proteases typically cannot directly drive transcription from a promoter, the linkage of protease activity to viral particle packaging is provided indirectly, for example, by using a transcriptional activator that is expressed in the host cells in an inactive form and a desired activity of the protease, e.g., protease-mediated cleavage, of the inactive form results in activation of the transcriptional activator and thus in an increase in the expression of the gene required for packaging infectious viral particles. In some embodiments, the transcriptional activator is expressed in the form of a fusion protein comprising an inhibitor of the transcriptional activator fused to the transcriptional activator via a linker comprising a protease cleavage site. Cleavage of the protease cleavage site results in a dissociation of the inhibitor from the transcriptional activator and thus a lowering of the effective concentration of inhibitor in the vicinity of the transcriptional activator. In some embodiments, cleavage of the linker by the protease of interest thus results in an increase in the transcriptional activity of the transcriptional activator, *e.g.*, by at least a factor of 2, 5, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10000, or more.

[00122] In some embodiments, the host cells comprise all phage genes except for the at least one gene for the generation of infectious phage particles in the form of a helper phage. In some embodiments, the phage genes on the helper phage include pI, pII, pIV, pV, pVI, pVII, pVIII, pIX, and/or pX. Some exemplary helper phages suitable for use in such embodiments are provided herein, and additional suitable helper phages will be apparent to the skilled artisan based on the instant disclosure. The host cell may also provide phage functions based on expression constructs other than helper phage, for example, expression constructs integrated into the host cell genome or provided on artificial chromosomes or on separate plasmids. One advantage of providing phage functions in the host cell, *e.g.*, by using a helper phage, is that the selection phage encoding the protease of interest can be deficient in genes encoding proteins or other functions provided by the host cell and can, accordingly, carry a longer gene encoding the protease of interest.

[00123] In some embodiments, the host cells comprise the accessory plasmid encoding the at least one gene for the generation of infectious phage particles, *e.g.*, of the M13 phage, encoding the protease to be evolved and a helper phage, and together, the helper phage and the accessory plasmid comprise all genes required for the generation of infectious phage particles. Accordingly, in some such embodiments, variants of the vector that do not encode a protease variant that can untether the inhibitor from the transcriptional activator will not efficiently be packaged, since they cannot effect an increase in expression of the gene required for the generation of infectious phage particles from the accessory plasmid. On the other hand, variants of the vector that encode a protease variant that can efficiently cleave the inhibitor from the transcriptional activator will effect increased transcription of the at least one gene required for the generation of infectious phage particles from the accessory plasmid and thus be efficiently packaged into infectious phage particles.

[00124] In some embodiments, diversifying the vector population is achieved by providing a flow of host cells that does not select for gain-of-function mutations in the gene encoding the protease of interest for replication, mutagenesis, and propagation of the population of vectors. In some embodiments, the host cells are host cells that express all genes required for the generation of infectious viral particles, for example, bacterial cells that express a complete helper phage, and, thus, do not impose selective pressure on the gene of interest. In other embodiments, the host cells comprise an accessory plasmid comprising a conditional promoter with a baseline activity sufficient to support viral vector propagation even in the absence of significant gain-of-function mutations of the gene of interest. This can be achieved by using a "leaky" conditional promoter, by using a high-copy number accessory

plasmid, thus amplifying baseline leakiness, and/or by using a conditional promoter on which the initial version of the gene of interest effects a low level of activity while a desired gain-of-function mutation effects a significantly higher activity.

[00125] Such methods involving host cells of varying selective stringency or varying the selection stringency in other ways as described herein allow for harnessing the power of continuous evolution methods as provided herein for the evolution of protease functions that are completely absent in the initial version of the protease of interest, for example, for the evolution of proteases that target sequences that the initial protease used at the outset of the respective PACE experiments does not recognize or cleave at all.

[00126] In some embodiments, the protease PACE methods provided herein further comprises a negative selection for undesired protease activity in addition to the positive selection for a desired protease activity. Such negative selection methods are useful, for example, in order to maintain protease specificity when increasing the cleavage efficiency of a protease directed towards a specific target site. This can avoid, for example, the evolution of proteases that show a generally increased protease activity, including an increased protease activity towards off-target sites, which is generally undesired in the context of therapeutic proteases.

[00127] In some embodiments, negative selection is applied during a continuous evolution process as described herein, by penalizing the undesired activities of evolved proteases. This is useful, for example, if the desired evolved protease is an enzyme with high specificity for a target site, for example, a protease with altered, but not broadened, specificity. In some embodiments, negative selection of an undesired activity, *e.g.*, off-target protease activity, is achieved by causing the undesired activity to interfere with pIII production, thus inhibiting the propagation of phage genomes encoding gene products with an undesired activity. In some embodiments, expression of a dominant-negative version of pIII or expression of an antisense RNA complementary to the gIII RBS and/or gIII start codon is linked to the presence of an undesired protease activity. Suitable negative selection strategies and reagents useful for negative selection, such as dominant-negative versions of M13 pIII, are described herein and in International PCT Application, PCT/US2011/066747, filed December 22, 2011, published as WO 2012/088381 on June 28, 2012; and U.S. Application, U.S.S.N. 13/922,812, filed June 20, 2013, the entire contents of each of which are incorporated herein by reference.

[00128] In some embodiments, counter-selection against activity on non-target substrates is achieved by linking undesired evolved protease activities to the inhibition of

phage propagation. In some embodiments, a dual selection strategy is applied during a continuous evolution experiment, in which both positive selection and negative selection constructs are present in the host cells. In some such embodiments, the positive and negative selection constructs are situated on the same plasmid, also referred to as a dual selection accessory plasmid.

[00129] One advantage of using a simultaneous dual selection strategy is that the selection stringency can be fine-tuned based on the activity or expression level of the negative selection construct as compared to the positive selection construct. Another advantage of a dual selection strategy is that the selection is not dependent on the presence or the absence of a desired or an undesired activity, but on the ratio of desired and undesired activities, and, thus, the resulting ratio of pIII and pIII-neg that is incorporated into the respective phage particle.

[00130] For example, in some embodiments, the host cells comprise an expression construct encoding a dominant-negative form of the at least one gene for the generation of infectious phage particles, *e.g.*, a dominant-negative form of the pIII protein (pIII-neg), under the control of an inducible promoter that is activated by a transcriptional activator other than the transcriptional activator driving the positive selection system. Expression of the dominant-negative form of the gene diminishes or completely negates any selective advantage an evolved phage may exhibit and thus dilutes or eradicates any variants exhibiting undesired activity from the lagoon.

[00131] For example, if the positive selection system comprises a T7 promoter driving the expression of the at least one gene for the generation of infectious phage particles, and a T7 RNA polymerase fused to a T7-RNA polymerase inhibitor via a linker comprising a protease target site that is cleaved by a desired protease activity, the negative selection system should be a non-T7 based system. For example, in some such embodiments, the negative selection system could be based on T3 polymerase activity, *e.g.*, in that it comprises a T3 promoter driving the expression of a dominant-negative form of the at least one gene for the generation of infectious phage particles, and a T3 RNA polymerase fused to a T3-RNA polymerase inhibitor via a linker comprising a protease target site that is cleaved by an undesired protease activity. When used together, such positive-negative PACE selection results in the evolution of proteases that exhibit the desired activity but not the undesired activity. In some embodiments, the undesired function is cleavage of an off-target protease cleavage site. In some embodiments, the undesired function is cleavage of the linker sequence of the fusion protein outside of the protease cleavage site.

[00132] Some aspects of this invention provide or utilize a dominant negative variant of pIII (pIII-neg). These aspects are based on the recognition that a pIII variant that comprises the two N-terminal domains of pIII and a truncated, termination-incompetent C-terminal domain is not only inactive but is a dominant-negative variant of pIII. A pIII variant comprising the two N-terminal domains of pIII and a truncated, termination-incompetent C-terminal domain was described in Bennett, N. J.; Rakonjac, J., Unlocking of the filamentous bacteriophage virion during infection is mediated by the C domain of pIII. *Journal of Molecular Biology* 2006, 356 (2), 266-73; the entire contents of which are incorporated herein by reference. The dominant negative property of such pIII variants has been described in more detail in PCT Application PCT/US2011/066747, published as WO2012/088381 on June 28, 2012, the entire contents of which are incorporated herein by reference.

[00133] The pIII-neg variant as provided in some embodiments herein is efficiently incorporated into phage particles, but it does not catalyze the unlocking of the particle for entry during infection, rendering the respective phage noninfectious even if wild type pIII is present in the same phage particle. Accordingly, such pIII-neg variants are useful for devising a negative selection strategy in the context of PACE, for example, by providing an expression construct comprising a nucleic acid sequence encoding a pIII-neg variant under the control of a promoter comprising a recognition motif, the recognition of which is undesired. In other embodiments, pIII-neg is used in a positive selection strategy, for example, by providing an expression construct in which a pIII-neg encoding sequence is controlled by a promoter comprising a nuclease target site or a repressor recognition site, the recognition of either one is desired.

[00134] In some embodiments, the vector or phage encoding the protease to be evolved is a filamentous phage, for example, an M13 phage, such as an M13 selection phage as described in more detail elsewhere herein. In some embodiments, the host cells are cells amenable to infection by the filamentous phage, *e.g.*, by M13 phage, such as, for example, *E. coli* cells. In some such embodiments, the gene required for the production of infectious viral particles is the M13 gene III (gIII) encoding the M13 protein III (pIII).

[00135] Typically, the vector/host cell combination is chosen in which the life cycle of the vector is significantly shorter than the average time between cell divisions of the host cell. Average cell division times and vector life cycle times are well known in the art for many cell types and vectors, allowing those of skill in the art to ascertain such host cell/vector combinations. In certain embodiments, host cells are being removed from the population of

host cells in which the vector replicates at a rate that results in the average time of a host cell remaining in the host cell population before being removed to be shorter than the average time between cell divisions of the host cells, but to be longer than the average life cycle of the viral vector employed. The result of this is that the host cells, on average, do not have sufficient time to proliferate during their time in the host cell population while the viral vectors do have sufficient time to infect a host cell, replicate in the host cell, and generate new viral particles during the time a host cell remains in the cell population. This assures that the only replicating nucleic acid in the host cell population is the vector encoding the protease to be evolved, and that the host cell genome, the accessory plasmid, or any other nucleic acid constructs cannot acquire mutations allowing for escape from the selective pressure imposed.

[00136] For example, in some embodiments, the average time a host cell remains in the host cell population is about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19, about 20, about 21, about 22, about 23, about 24, about 25, about 30, about 35, about 40, about 45, about 50, about 55, about 60, about 70, about 80, about 90, about 100, about 120, about 150, or about 180 minutes.

[00137] In some embodiments, the average time a host cell remains in the host cell population depends on how fast the host cells divide and how long infection (or conjugation) requires. In general, the flow rate should be faster than the average time required for cell division, but slow enough to allow viral (or conjugative) propagation. The former will vary, for example, with the media type, and can be delayed by adding cell division inhibitor antibiotics (FtsZ inhibitors in *E. coli*, *etc.*). Since the limiting step in continuous evolution is production of the protein required for gene transfer from cell to cell, the flow rate at which the vector washes out will depend on the current activity of the gene(s) of interest. In some embodiments, titrable production of the protein required for the generation of infectious particles, as described herein, can mitigate this problem. In some embodiments, an indicator of phage infection allows computer-controlled optimization of the flow rate for the current activity level in real-time.

[00138] In some embodiments, a protease PACE experiment according to methods provided herein is run for a time sufficient for at least 10, at least 20, at least 30, at least 40, at least 50, at least 100, at least 200, at least 300, at least 400, at least, 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1250, at least 1500, at least 1750, at least 2000, at least 2500, at least 3000, at least 4000, at least 5000, at least 7500, at least 10000, or more consecutive viral life cycles. In certain embodiments, the viral vector is an M13 phage, and the length of a single viral life cycle is about 10-20 minutes.

[00139] In some embodiments, the host cells are contacted with the vector and/or incubated in suspension culture. For example, in some embodiments, bacterial cells are incubated in suspension culture in liquid culture media. Suitable culture media for bacterial suspension culture will be apparent to those of skill in the art, and the invention is not limited in this regard. See, for example, *Molecular Cloning: A Laboratory Manual*, 2nd Ed., ed. by Sambrook, Fritsch, and Maniatis (Cold Spring Harbor Laboratory Press: 1989); Elizabeth Kutter and Alexander Sulakvelidze: *Bacteriophages: Biology and Applications*. CRC Press; 1st edition (December 2004), ISBN: 0849313368; Martha R. J. Clokie and Andrew M. Kropinski: *Bacteriophages: Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions* (*Methods in Molecular Biology*) Humana Press; 1st edition (December, 2008), ISBN: 1588296822; Martha R. J. Clokie and Andrew M. Kropinski: *Bacteriophages: Methods and Protocols, Volume 2: Molecular and Applied Aspects* (*Methods in Molecular Biology)* Humana Press; 1st edition (December 2008), ISBN: 1603275649; all of which are incorporated herein in their entirety by reference for disclosure of suitable culture media for bacterial host cell culture).

[00140] Suspension culture typically requires the culture media to be agitated, either continuously or intermittently. This is achieved, in some embodiments, by agitating or stirring the vessel comprising the host cell population. In some embodiments, the outflow of host cells and the inflow of fresh host cells is sufficient to maintain the host cells in suspension. This in particular, if the flow rate of cells into and/or out of the lagoon is high.

[00141] In some embodiments, the flow of cells through the lagoon is regulated to result in an essentially constant number of host cells within the lagoon. In some embodiments, the flow of cells through the lagoon is regulated to result in an essentially constant number of fresh host cells within the lagoon. Typically, the lagoon will hold host cells in liquid media, for example, cells in suspension in a culture media. However, lagoons in which adherent host cells are cultured on a solid support, such as on beads, membranes, or appropriate cell culture surfaces are also envisioned. The lagoon may comprise additional features, such as a stirrer or agitator for stirring or agitating the culture media, a cell densitometer for measuring cell density in the lagoon, one or more pumps for pumping fresh host cells into the culture vessel and/or for removing host cells from the culture vessel, a thermometer and/or thermocontroller for adjusting the culture temperature, as well as sensors for measuring pH, osmolarity, oxygenation, and other parameters of the culture media. The lagoon may also comprise an inflow connected to a holding vessel comprising a mutagen or a transcriptional inducer of a conditional gene expression system, such as the arabinose-

inducible expression system of the mutagenesis plasmid described in more detail elsewhere herein.

[00142] In some embodiments, the host cell population is continuously replenished with fresh, uninfected host cells. In some embodiments, this is accomplished by a steady stream of fresh host cells into the population of host cells. In other embodiments, however, the inflow of fresh host cells into the lagoon is semi-continuous or intermittent (*e.g.,* batch-fed). In some embodiments, the rate of fresh host cell inflow into the cell population is such that the rate of removal of cells from the host cell population is compensated. In some embodiments, the result of this cell flow compensation is that the number of cells in the cell population is substantially constant over the time of the continuous evolution procedure. In some embodiments, the portion of fresh, uninfected cells in the cell population is substantially constant over the time of the continuous evolution procedure. For example, in some embodiments, about 10%, about 15%, about 20%, about 25%, about 30%, about 40%, about 50%, about 60%, about 70%, about 75%, about 80%, or about 90% of the cells in the host cell population are not infected by virus. In general, the faster the flow rate of host cells is, the smaller the portion of cells in the host cell population that are infected will be. However, faster flow rates allow for more transfer cycles, *e.g.*, viral life cycles, and, thus, for more generations of evolved vectors in a given period of time, while slower flow rates result in a larger portion of infected host cells in the host cell population and therefore a larger library size at the cost of slower evolution. In some embodiments, the range of effective flow rates is invariably bounded by the cell division time on the slow end and vector washout on the high end In some embodiments, the viral load, for example, as measured in infectious viral particles per volume of cell culture media is substantially constant over the time of the continuous evolution procedure.

[00143] Typically, the fresh host cells introduced into the lagoon comprise the protease selection system, *e.g.*, the accessory plasmid encoding the at least one gene for the generation of infectious phage particles, the expression construct encoding the fusion protein of the transcriptional activator fused to an inhibitor via a protease-cleavable linker, and an expression construct providing other phage functions, such as, for example, a helper phage. In some embodiments, the host cells also comprise a mutagenesis plasmid. In some embodiments, the host cells are generated by contacting an uninfected host cell with the relevant vectors, for example, the accessory plasmid, a construct encoding the protease-cleavable fusion protein, the helper phage, and the mutagenesis plasmid, and growing an amount of host cells sufficient for the replenishment of the host cell population in a

continuous evolution experiment. Methods for the introduction of plasmids and other gene constructs into host cells are well known to those of skill in the art and the invention is not limited in this respect. For bacterial host cells, such methods include, but are not limited to electroporation and heat-shock of competent cells. In some embodiments, the accessory plasmid comprises a selection marker, for example, an antibiotic resistance marker, and the fresh host cells are grown in the presence of the respective antibiotic to ensure the presence of the plasmid in the host cells. Where multiple plasmids are present, different markers are typically used. Such selection markers and their use in cell culture are known to those of skill in the art, and the invention is not limited in this respect.

[00144] In some embodiments, the host cell population in a continuous evolution experiment is replenished with fresh host cells growing in a parallel, continuous culture. In some embodiments, the cell density of the host cells in the host cell population contacted with the viral vector and the density of the fresh host cell population is substantially the same.

[00145] Typically, the cells being removed from the cell population contacted with the vector comprise cells that are infected with the vector and uninfected cells. In some embodiments, cells are being removed from the cell populations continuously, for example, by effecting a continuous outflow of the cells from the population. In other embodiments, cells are removed semi-continuously or intermittently from the population. In some embodiments, the replenishment of fresh cells will match the mode of removal of cells from the cell population, for example, if cells are continuously removed, fresh cells will be continuously introduced. However, in some embodiments, the modes of replenishment and removal may be mismatched, for example, a cell population may be continuously replenished with fresh cells, and cells may be removed semi-continuously or in batches.

[00146] In some embodiments, the rate of fresh host cell replenishment and/or the rate of host cell removal is adjusted based on quantifying the host cells in the cell population. For example, in some embodiments, the turbidity of culture media comprising the host cell population is monitored and, if the turbidity falls below a threshold level, the ratio of host cell inflow to host cell outflow is adjusted to effect an increase in the number of host cells in the population, as manifested by increased cell culture turbidity. In other embodiments, if the turbidity rises above a threshold level, the ratio of host cell inflow to host cell outflow is adjusted to effect a decrease in the number of host cells in the population, as manifested by decreased cell culture turbidity. Maintaining the density of host cells in the host cell population within a specific density range ensures that enough host cells are available as hosts for the evolving viral vector population, and avoids the depletion of nutrients at the cost

of viral packaging and the accumulation of cell-originated toxins from overcrowding the culture.

[00147] In some embodiments, the cell density in the host cell population and/or the fresh host cell density in the inflow is about 10^2 cells/ml to about 10^{12} cells/ml. In some embodiments, the host cell density is about 10^2 cells/ml, about 10^3 cells/ml, about 10^4 cells/ml, about 10^5 cells/ml, about $5 \cdot 10^5$ cells/ml, about 10^6 cells/ml, about $5 \cdot 10^6$ cells/ml, about 10^7 cells/ml, about $5 \cdot 10^7$ cells/ml, about 10^8 cells/ml, about $5 \cdot 10^8$ cells/ml, about 10^9 cells/ml, about $5 \cdot 10^9$ cells/ml, about 10^{10} cells/ml, or about $5 \cdot 10^{10}$ cells/ml. In some embodiments, the host cell density is more than about 10^{10} cells/ml.

[00148] The protease PACE methods provided herein are typically carried out in a lagoon. Suitable lagoons and other laboratory equipment for carrying out protease PACE methods as provided herein have been described in detail elsewhere. See, for example, International PCT Application, PCT/US2011/066747, published as WO2012/088381 on June 28, 2012, the entire contents of which are incorporated herein by reference. In some embodiments, the lagoon comprises a cell culture vessel comprising an actively replicating population of vectors, for example, phage vectors comprising a gene encoding the protease of interest, and a population of host cells, for example, bacterial host cells. In some embodiments, the lagoon comprises an inflow for the introduction of fresh host cells into the lagoon and an outflow for the removal of host cells from the lagoon. In some embodiments, the inflow is connected to a turbidostat comprising a culture of fresh host cells. In some embodiments, the outflow is connected to a waste vessel, or a sink. In some embodiments, the lagoon further comprises an inflow for the introduction of a mutagen into the lagoon. In some embodiments that inflow is connected to a vessel holding a solution of the mutagen. In some embodiments, the lagoon comprises an inflow for the introduction of an inducer of gene expression into the lagoon, for example, of an inducer activating an inducible promoter within the host cells that drives expression of a gene promoting mutagenesis (*e.g.*, as part of a mutagenesis plasmid), as described in more detail elsewhere herein. In some embodiments, that inflow is connected to a vessel comprising a solution of the inducer, for example, a solution of arabinose.

[00149] In some embodiments, the lagoon comprises a controller for regulation of the inflow and outflow rates of the host cells, the inflow of the mutagen, and/or the inflow of the inducer. In some embodiments, a visual indicator of phage presence , for example, a fluorescent marker, is tracked and used to govern the flow rate, keeping the total infected population constant. In some embodiments, the visual marker is a fluorescent protein

encoded by the phage genome, or an enzyme encoded by the phage genome that, once expressed in the host cells, results in a visually detectable change in the host cells. In some embodiments, the visual tracking of infected cells is used to adjust a flow rate to keep the system flowing as fast as possible without risk of vector washout.

[00150] In some embodiments, the controller regulates the rate of inflow of fresh host cells into the lagoon to be substantially the same (volume/volume) as the rate of outflow from the lagoon. In some embodiments, the rate of inflow of fresh host cells into and/or the rate of outflow of host cells from the lagoon is regulated to be substantially constant over the time of a continuous evolution experiment. In some embodiments, the rate of inflow and/or the rate of outflow is from about 0.1 lagoon volumes per hour to about 25 lagoon volumes per hour. In some embodiments, the rate of inflow and/or the rate of outflow is approximately 0.1 lagoon volumes per hour (lv/h), approximately 0.2 lv/h, approximately 0.25 lv/h, approximately 0.3 lv/h, approximately 0.4 lv/h, approximately 0.5 lv/h, approximately 0.6 lv/h, approximately 0.7 lv/h, approximately 0.75 lv/h, approximately 0.8 lv/h, approximately 0.9 lv/h, approximately 1 lv/h, approximately 2 lv/h, approximately 2.5 lv/h, approximately 3 lv/h, approximately 4 lv/h, approximately 5 lv/h, approximately 7.5 lv/h, approximately 10 lv/h, or more than 10 lv/h.

[00151] In some embodiments, the inflow and outflow rates are controlled based on a quantitative assessment of the population of host cells in the lagoon, for example, by measuring the cell number, cell density, wet biomass weight per volume, turbidity, or cell growth rate. In some embodiments, the lagoon inflow and/or outflow rate is controlled to maintain a host cell density of from about 10^2 cells/ml to about 10^{12} cells/ml in the lagoon. In some embodiments, the inflow and/or outflow rate is controlled to maintain a host cell density of about 10^2 cells/ml, about 10^3 cells/ml, about 10^4 cells/ml, about 10^5 cells/ml, about $5x10^5$ cells/ml, about 10^6 cells/ml, about $5x10^6$ cells/ml, about 10^7 cells/ml, about $5x10^7$ cells/ml, about 10^8 cells/ml, about $5x10^8$ cells/ml, about 10^9 cells/ml, about $5x10^9$ cells/ml, about 10^{10} cells/ml, about $5x10^{10}$ cells/ml, or more than $5x10^{10}$ cells/ml, in the lagoon. In some embodiments, the density of fresh host cells in the turbidostat and the density of host cells in the lagoon are substantially identical.

[00152] In some embodiments, the lagoon inflow and outflow rates are controlled to maintain a substantially constant number of host cells in the lagoon. In some embodiments, the inflow and outflow rates are controlled to maintain a substantially constant frequency of fresh host cells in the lagoon. In some embodiments, the population of host cells is continuously replenished with fresh host cells that are not infected by the phage. In some

embodiments, the replenishment is semi-continuous or by batch-feeding fresh cells into the cell population.

[00153] In some embodiments, the lagoon volume is from approximately 1ml to approximately 100 l, for example, the lagoon volume is approximately 1 ml, approximately 10 ml, approximately 50 ml, approximately 100 ml, approximately 200 ml, approximately 250 ml, approximately 500ml, approximately 750ml, approximately 1 l, approximately 2ml, approximately 2.5l, approximately 3l, approximately 4l, approximately 5l, approximately 10l, approximately 1ml-10ml, approximately 10ml-50ml, approximately 50ml-100, approximately 100ml-250ml, approximately 250ml-500ml, approximately 500ml-1l, approximately 1l-2l, approximately 2l-5l, approximately 5l-10l,approximately 10-50l, approximately 50-100l, or more than 100l.

[00154] In some embodiments, the lagoon and/or the turbidostat further comprises a heater and a thermostat controlling the temperature. In some embodiments, the temperature in the lagoon and/or the turbidostat is controlled to be from about 4 °C to about 55 °C, preferably from about 25 °C to about 39 °C, for example, about 37 °C.

[00155] In some embodiments, the inflow rate and/or the outflow rate is controlled to allow for the incubation and replenishment of the population of host cells for a time sufficient for at least 10, at least 20, at least 30, at least 40, at least 50, at least 100, at least 200, at least 300, at least 400, at least, 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1250, at least 1500, at least 1750, at least 2000, at least 2500, at least 3000, at least 4000, at least 5000, at least 7500, at least 10000, or more consecutive vector or phage life cycles. In some embodiments, the time sufficient for one phage life cycle is about 10, 15, 20, 25, or 30 minutes.

[00156] Therefore, in some embodiments, the time of the entire evolution procedure is about 12 hours, about 18 hours, about 24 hours, about 36 hours, about 48 hours, about 50 hours, about 3 days, about 4 days, about 5 days, about 6 days, about 7 days, about 10 days, about two weeks, about 3 weeks, about 4 weeks, or about 5 weeks.

[00157] In some embodiments, a PACE method as provided herein is performed in a suitable apparatus as described herein. For example, in some embodiments, the apparatus comprises a lagoon that is connected to a turbidostat comprising a host cell as described herein. In some embodiments, the host cell is an *E. coli* host cell. In some embodiments, the host cell comprises an accessory plasmid as described herein, a helper plasmid as described herein, a mutagenesis plasmid as described herein, and/or an expression construct encoding a fusion protein as described herein, or any combination thereof. In some embodiments, the

lagoon further comprises a selection phage as described herein, for example, a selection phage encoding a protease of interest. In some embodiments, the lagoon is connected to a vessel comprising an inducer for a mutagenesis plasmid, for example, arabinose. In some embodiments, the host cells are *E. coli* cells comprising the F' plasmid, for example, cells of the genotype F′proA$^+$B$^+$ Δ(lacIZY) zzf::Tn10(TetR)/ endA1 recA1 galE15 galK16 nupG rpsL ΔlacIZYA araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) proBA::pir116 λ$^-$.

[00158] For example, in some embodiments, a PACE method as provided herein is carried out in an apparatus comprising a lagoon of about 100 ml, or about 1 l volume, wherein the lagoon is connected to a turbidostat of about 0.5 l, 1 l , or 3 l volume, and to a vessel comprising an inducer for a mutagenesis plasmid, for example, arabinose, wherein the lagoon and the turbidostat comprise a suspension culture of *E. coli* cells at a concentration of about 5 x 10^8 cells/ml. In some embodiments, the flow of cells through the lagoon is regulated to about 3 lagoon volumes per hour. In some embodiments, cells are removed from the lagoon by continuous pumping, for example, by using a waste needle set at a height of the lagoon vessel that corresponds to a desired volume of fluid (*e.g.*, about 100 ml, in the lagoon. In some embodiments, the host cells are *E. coli* cells comprising the F' plasmid, for example, cells of the genotype F′proA$^+$B$^+$ Δ(lacIZY) zzf::Tn10(TetR)/ endA1 recA1 galE15 galK16 nupG rpsL ΔlacIZYA araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) proBA::pir116 λ$^-$.

Evaluation and selection of protease inhibitors

[00159] Some aspects of this disclosure provide method for generating protease variants that are resistant to a protease inhibitor. Such methods are useful, for example, in the evaluation of the likelihood of encountering inhibitor-resistant protease variants after administration of an inhibitor to a subject or a population of subjects. In the context of therapeutic protease inhibitors, the occurrence of resistant protease variants may render a protease inhibitor drug ineffective, as seen with some of the antiviral protease inhibitors described herein. The protease PACE methods provided herein allow for an evaluation of the occurrence of inhibitor-resistant protease variants *in vitro*, *e.g.*, before a drug candidate is tested in a clinical setting, thus contributing to saving time and expense during drug development and avoiding the development of protease inhibitor drugs that later become ineffective.

[00160] In some embodiments, a protease inhibitor inhibits the activity of its target protease in a dose-dependent manner. In some embodiments, the inhibitory potential of a

protease inhibitor referred to herein is its maximum level of inhibition or its maximum level of inhibition at a non-toxic concentration. In some embodiments, a protease inhibitor inhibits the activity of a target protease by at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, at least 97.5%, at least 98%, at least 99%, at least 99.5%, at least 99.9%, or by at least 99.99%. In some embodiments, a protease inhibitor inhibits the activity of a protease below measurable levels.

[00161] In some embodiments, the protease PACE methods for evaluating and selecting protease inhibitors are similar to the protease PACE methods provided elsewhere herein, with the main difference being that they are carried out in the presence of a protease inhibitor. Accordingly, in some embodiments, such methods comprise (a) contacting a population of host cells with a population of phage vectors comprising a gene encoding a protease and deficient in at least one gene for the generation of infectious phage particles, wherein (1) the host cells are amenable to transfer of the vector; (2) the vector allows for expression of the protease in the host cell, can be replicated by the host cell, and the replicated vector can transfer into a second host cell; (3) the host cell expresses a gene product encoded by the at least one gene for the generation of infectious phage particles of (a) in response to the activity of the protease, and the level of gene product expression depends on the activity of the protease; (b) incubating the population of host cells in the presence of a protease inhibitor and under conditions allowing for mutation of the gene encoding the protease and the transfer of the vector comprising the gene encoding the protease of interest from host cell to host cell, wherein host cells are removed from the host cell population, and the population of host cells is replenished with fresh host cells that do not harbor the vector; and (c) isolating a replicated vector from the host cell population of step (b). The replicated vector isolated in step (c) thus encodes a mutated protease variant that exhibits protease activity in the presence of the protease inhibitor and/or increased protease activity in the presence of the inhibitor as compared to the original version of the protease..

[00162] In some embodiments, the method comprises analyzing the mutation(s) present in a plurality of protease variants isolated in step (c), and comparing the observed mutations to identify shared and non-shared mutations. In some embodiments, the shared mutations are identified as resistance-conferring mutations, and in some embodiments, the non-shared mutations are identified as background mutations that do confer resistance by themselves or do not contribute to resistance at all. In some embodiments, the method further comprises confirming the identification of a mutation identified to confer resistance, for example, by generating a protease variant based on the original protease version by

introducing only the mutation identified as conferring resistance and then measuring the effect of the mutation on the protease activity of the protease of interest.

[00163] In some embodiments, the method further comprises testing a plurality of individual protease inhibitors in separate experiments under identical conditions, and identifying the protease inhibitor to which no inhibitor-resistant protease variant could be identified. While a negative PACE experiment with a protease inhibitor as provided herein does not guarantee that no inhibitor-resistant protease variants will form after an inhibitor is used in a clinical setting, such an observation provides evidence that the range of mutations that are accessible in a PACE experiment are not sufficient to generate a resistant protease variant. With other parameters, such as toxicity and pharmacodynamics parameters, being equal or similar to other inhibitor candidates that yield inhibitor-resistant protease variants in the PACE experiments provided herein, such a "PACE-negative" inhibitor candidate constitutes a preferable lead compound, since the likelihood of the emergence of inhibitor-resistant proteases in the clinic is lower as compared to the "PACE-positive" comparison candidates.

[00164] In some embodiments, the method further comprises testing a plurality of individual protease inhibitors in separate experiments under identical conditions, and identifying the protease inhibitor that required the highest number of mutations in the protease to create an inhibitor-resistant protease variant. Such methods are useful, for example, if during a comparison of inhibitor candidates no candidate for which no inhibitor-resistant protease variants could be generated was identified. In the clinical context, it is less likely for an inhibitor-resistant protease to emerge, if multiple mutations are required in order for the protease to gain resistance as compared to proteases that gain resistance from a single point mutation.

[00165] In some embodiments, the method may be repeated multiple times, for example, in that a first round of protease PACE experiments is conducted in which a number of candidate inhibitors is compared. At the end of the first round, a candidate inhibitor with a low or with the lowest likelihood of the emergence of resistant proteases is selected, *e.g.*, a candidate inhibitor that did not yield resistant protease variants in the PACE experiment, or that yielded resistant proteases requiring the highest number of mutations as compared to the other candidate inhibitors. A plurality of candidate inhibitor variants is then generated based on the general structure of the selected candidate from the first round, and the candidate variants are subjected to a round of protease PACE. The candidate variant with no incidence of resistant protease variants or with the highest number of required mutations for the

emergence of resistant proteases is then selected. This process may be repeated several times, resulting in a more and more refined and optimized version of the inhibitor candidate. As the emergence of resistant protease variants becomes less and less likely in subsequent round, the stringency of selection may be reduced in order to still be able to measure further improvements in the design of the inhibitor.

[00166] In some embodiments, the host cells, accessory plasmids, lagoons, flow rates, culture conditions, mutagens, etc., can be chosen for the methods provided herein as described in more detail for other protease PACE methods herein.

Fusion proteins

[00167] One important aspect of this disclosure is the provision of fusion proteins that link protease activity to transcriptional activity and thus translate protease activity into regulation of gene expression and thus phage packaging efficiency in the context of protease PACE. Typically, this link is provided by a fusion protein comprising a domain able to activate transcription from a target promoter and an inhibitory domain that inhibits this transcriptional activity. The two domains are fused via a linker comprising a protease target site that is cleaved by a desired protease activity to be evolved, and the severance of the inhibitory domain from the transcriptional activator domain relieves the inhibition of the activator, which in turn can drive expression from the target promoter.

[00168] Accordingly, some aspects of this disclosure provide fusion proteins comprising (a) a transcriptional activator; and (b) an inhibitor of the transcriptional activator of (a), wherein the inhibitor is fused to the transcriptional activator of (a) via a linker comprising a protease cleavage site. In some embodiments, the inhibitor of (ii) is fused to the N-terminus of the transcriptional activator of (i). In some embodiments, the inhibitor of (ii) is fused to the C-terminus of the transcriptional activator of (i). Some suitable pairs of transcriptional activator and inhibitor for the generation of fusion proteins according to some aspects of this disclosure are provided herein and additional suitable pairs and fusion proteins comprising such pairs will be apparent to those of skill in the art based on the instant disclosure. In some embodiments, the inhibitor may be a protein degradation tag that leads to rapid degradation of the tagged protein (e.g., the transcriptional activator) unless the tag is cleaved off by a protease. In other embodiments, the inhibitor may be a signal peptide that targets the fusion protein for export out of the cell or into a cellular compartment in which the transcriptional activator cannot activate transcription of the target gene. Upon cleavage of the signal peptide from the transcriptional activator by a protease, the untagged activator is not

exported from the cell and can activate transcription from the target promoter. Suitable signal peptides include, without limitation, periplasmic export tags, e.g., those from PelB, OmpA, PhoA, pIII, and other exported proteins known to those of skill in the art. Suitable transcriptional activators include, without limitation, RNA polymerases, e.g., from bacteriophages other than T7 (e.g., T3, T4, etc.) and from other organisms, as well as sequence specific transcriptional activators such as, for example, Gal4.

[00169] In some embodiments, the transcriptional activity of the fusion protein is inhibited as compared to the activity of the transcriptional activator alone. In some embodiments, the transcriptional activity of the fusion protein is less than 50%, less than 25%, less than 20%, less than 10%, less than 5%, less than 4%, less than 3%, less than 1%, less than 0.1%, less than 0.01%, or less than 0.001% the activity of the transcriptional activator alone.

[00170] In some embodiments, cleavage of the protease cleavage site results in an activation of the transcriptional activator. In some embodiments, the transcriptional activator regains 100% of the transcriptional activity upon cleavage of the protease cleavage site as compared to the transcriptional activator alone (*e.g.*, in its original form not fused to an inhibitor). In some embodiments, the activity of the transcriptional activator comprised in the fusion protein is increased at least2-fold, at least5-fold, at least10-fold, at least20-fold, at least25-fold, at least 50-fold, at least 75-fold, at least 100-fold, at least 200-foled, at least 250-fold, at least 500-fold, at least 750-fold, at least 1000-fold, at least2000-fold, at least2500-fold, at least 5000-fold, at least 7500-fold, or at least 10000-fold upon cleavage of the protease cleavage site of the linker.

[00171] Those of skill in the art will be able to readily envision suitable methods and assays for determining the transcriptional activity of a given protein, *e.g.*, of a transcriptional activator bound to an inhibitor or alone, *e.g.*, after cleavage of a fusion protein as provided herein. Such assays typically include the use of a reporter construct harboring a promoter targeted by the transcriptional activator (*e.g.*, a T7 promoter if the transcriptional activator is a T7 RNA polymerase) and a reporter gene, *e.g.*, a gene encoding a fluorescent protein or an enzyme catalyzing a bioluminescent reaction. Some exemplary methods and assays are provided herein and additional suitable methods and assays will be apparent to those of skill in the art based on the instant disclosure.

[00172] In some embodiments, the linker connecting the transcriptional activator and the inhibitor comprises a stretch of small residues, such as, for example, glycine, serine and/or alanine. Without wishing to be bound by theory, it is presumed that such stretches are

highly flexible and thus provide the flexibility required in some embodiments, e.g., in embodiments, in which a linker connects a transcriptional activator and an inhibitor that binds the transcriptional activator at a specific site or in a specific orientation. In some embodiments, a stretch of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, or at least 20 small residues (e.g., glycine, serine, and/or alanine, or any combination thereof), is comprised in the linker sequence, for example, on one or on both sides of the protease cleavage sequence.

Vectors and Reagents

[00173] Some aspects of this disclosure provide vectors and reagents for carrying out the inventive continuous protease evolution processes.

[00174] In some embodiments, a selection phage is provided that comprises a phage genome deficient in at least one gene required for the generation of infectious phage particles and a gene encoding a protease of interest to be evolved.

[00175] For example, in some embodiments, a selection phage as described in in PCT Application PCT/US2009/056194, published as WO2010/028347 on March 11, 2010; PCT Application PCT/US2011/066747, published as WO2012/088381 on June 28, 2012; and U.S. Nonprovisional Application 13/922,812, filed on June 20/2013, the entire contents of each of which are incorporated herein by reference, is provided, that comprises a multiple cloning site for insertion of a nucleic acid sequence encoding a protease of interest.

[00176] Such selection phage vectors typically comprise an M13 phage genome deficient in a gene required for the generation of infectious M13 phage particles, for example, a full-length gIII. In some embodiments, the selection phage comprises a phage genome providing all other phage functions required for the phage life cycle except the gene required for generation of infectious phage particles. In some such embodiments, an M13 selection phage is provided that comprises a gI, gII, gIV, gV, gVI, gVII, gVIII, gIX, and a gX gene, but not a full-length gIII. In some embodiments, the selection phage comprises a 3′-fragment of gIII, but no full-length gIII. The 3′-end of gIII comprises a promoter and retaining this promoter activity is beneficial, in some embodiments, for an increased expression of gVI, which is immediately downstream of the gIII 3′-promoter, or a more balanced (wild-type phage-like) ratio of expression levels of the phage genes in the host cell, which, in turn, can lead to more efficient phage production. In some embodiments, the 3′-fragment of gIII gene

comprises the 3′-gIII promoter sequence. In some embodiments, the 3′-fragment of gIII comprises the last 180 bp, the last 150 bp, the last 125 bp, the last 100 bp, the last 50 bp, or the last 25 bp of gIII. In some embodiments, the 3′- fragment of gIII comprises the last 180 bp of gIII. In some embodiments, the multiple cloning site for insertion of the gene encoding the protease of interest is located downstream of the gVIII 3′-terminator and upstream of the gIII-3′-promoter.

[00177] Some aspects of this disclosure provide nucleic acid constructs encoding a fusion protein as provided herein. For example, some aspects of this disclosure provide nucleic acid constructs comprising (a) a nucleic acid sequence encoding a transcriptional activator; (b) a nucleic acid sequence encoding an inhibitor of the of the transcriptional activator of (a); and (c) a nucleic acid sequence separating the nucleic acid sequences of (a) and (b), wherein the nucleic acid sequence of (c) encodes a linker and comprises a multiple cloning site allowing for the insertion of a nucleic acid sequence encoding a protease cleavage site. In some embodiments, the nucleic acid sequence of (c) further encodes a linker connecting the transcriptional activator and the inhibitor. In some embodiments, the linker comprises a stretch of at least two consecutive glycine residues on each side of the protease cleavage site.

[00178] Some aspects of this invention provide a vector system for continuous evolution procedures, comprising of a viral vector, for example, a selection phage, a nucleic acid encoding a fusion protein as provided herein, and a matching accessory plasmid. In some embodiments, a vector system for phage-based continuous directed evolution is provided that comprises (a) a selection phage comprising a multiple cloning site for insertion of a gene encoding a protease of interest to be evolved, wherein the phage genome is deficient in at least one gene required to generate infectious phage; (b) a nucleic acid construct comprising a nucleic acid sequence encoding a transcriptional activator fused to an inhibitor of the of the transcriptional activator via a linker sequence and comprising a multiple cloning site for insertion of a nucleic acid sequence encoding a protease cleavage site into the linker sequence; and (c) an accessory plasmid comprising the at least one gene required to generate infectious phage particle under the control of a conditional promoter that can be activated by the transcriptional activator upon cleavage of the protease cleavage site.

[00179] In some embodiments, the selection phage is an M13 phage as described herein. For example, in some embodiments, the selection phage comprises an M13 genome including all genes required for the generation of phage particles, for example, gI, gII, gIV, gV, gVI, gVII, gVIII, gIX, and gX gene, but not a full-length gIII gene. In some

embodiments, the selection phage genome comprises an F1 or an M13 origin of replication. In some embodiments, the selection phage genome comprises a 3′-fragment of gIII gene. In some embodiments, the selection phage comprises a multiple cloning site upstream of the gIII 3′-promoter and downstream of the gVIII 3′-terminator for insertion of a gene encoding a protease of interest.

[00180] The vector system may further comprise a helper phage, wherein the selection phage does not comprise all genes for the generation of infectious phage particles, and wherein the helper phage complements the genome of the selection phage, so that the helper phage genome and the selection phage genome together comprise at least one functional copy of all genes for the generation of phage particles, but are deficient in at least one gene required for the generation of infectious phage particles, which is provided by an accessory plasmid.

[00181] In some embodiments, the vector system further comprises a mutagenesis plasmid, for example, an arabinose-inducible mutagenesis plasmid as described herein. In some embodiments, the mutagenesis plasmid comprises a gene expression cassette encoding a component of *E. coli* translesion synthesis polymerase V, a deoxyadenosine methylase, and/or a hemimethylated-GATC binding domain, or any combination thereof. In some embodiments, the component of E. coli translesion synthesis polymerase V is *umuC*. In some embodiments, the deoxyadenosine methylase is *dam*. In some embodiments, the hemimethylated-GATC binding domain is *seqA*.

Evolved proteases and used thereof

[00182] Some aspects of this invention provide evolved protease variants produced via the PACE methods and systems described herein. In some embodiments, the evolved protease variants exhibit a desirable protease activity, for example, in that they cleave their original target site with higher efficiency, in that they cleave an altered cleavage site, and/or in that they do not cleave an undesirable off-target site. In addition, protease variants that are resistant to protease inhibitors are provided herein. Such resistant variants are useful in determining whether a given candidate protease inhibitor is a viable candidate for clinical development or should be discarded or modified because of a high likelihood of inhibitor-resistant proteases emerging during clinical use.

[00183] For example, some aspects of this invention provide evolved tobacco etch virus (TEV) protease variants, HCV protease variants, and human *rhinovirus*-14 3C (HRV) protease variants, *e.g.*, variants comprising at least one of the mutations described herein. In

some embodiments, the variants provided herein comprise at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, or at least 10 mutations as described herein. In some embodiments, the term "protease variant" refers to a non-naturally occurring protease, *i.e.*, comprising an amino acid sequence that is not found in nature. Such non-naturally occurring protease variants are also referred to herein as recombinant proteases.

[00184] In some embodiments, a protease variant provided herein is a recombinant (not naturally occurring) protease variant that is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to a naturally-occurring, wild-type protease, and comprises at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9 or at least 10 mutations as described herein.

[00185] Naturally occurring, wild-type protease sequences that can serve as a basis or a starting point for the directed evolution strategies described herein or for the generation of protease variants described herein are well known to those of ordinary skill in the art.

[00186] For example, in some embodiments, a suitable TEV protease sequence comprises the following sequence:

GESLFKGPRD YNPISSTICH LTNESDGHTT SLYGIGFGPF IITNKHLFRR
NNGTLLVQSL HGVFKVKNTT TLQQHLIDGR DMIIIRMPKD FPPFPQKLKF
REPQREERIC LVTTNFQTKS MSSMVSDTSC TFPSSDGIFW KHWIQTKDGQ
CGSPLVSTRD GFIVGIHSAS NFTNTNNYFT SVPKNFMELL TNQEAQQWVS
GWRLNADSVL WGGHKVFMSK PEEPFQPVKE ATQLMNELVY SQ

(SEQ ID NO: 12).

[00187] For example, in some embodiments, a suitable TEV protease sequence comprises the following sequence:

APITAYAQQT RGLLGCIITS LTGRDKNQVE GEVQVVSTAT QSFLASCVNG
VCWTVFHGAG SKTLAGPKGP VTQMYTNVDQ DLVGWPAPPG ARSLTPCTCG
SSDLYLVTRH ADVIPVRRRG DSRGALLSPR PVSYLKGSSG GPLLCPSGHA
VGIFRAAVCT RGVAKAVDFI PVESMETTMR SP

(SEQ ID NO: 13).

[00188] For example, in some embodiments, a suitable human *rhinovirus*-14 3C (HRV) protease variant protease sequence comprises the following sequence:

GPNTEFALSL LRKNIMTITT SKGEFTGLGI HDRVCVIPTH AQPGDDVLVN
GQKIRVKDKY KLVDPENINL ELTVLTLDRN EKFRDIRGFI SEDLEGVDAT

LVVHSNNFTN TILEVGPVTM AGLINLSSTP TNRMIRYDYA TKTGQCGGVL
CATGKIFGIH VGGNGRQGFS AQLKKQYFVE KQ
(SEQ ID NO: 14).

[00189] Those of ordinary skill in the art will understand that the sequences provided here are exemplary and not meant to limit the scope of this disclosure. Additional suitable protease sequences will be apparent to those of ordinary skill in the art.

[00190] Some aspects of this invention provide methods and strategies to use proteases that have been evolved via the PACE methods provided herein to identify protease inhibitors. For example, in some embodiments, a therapeutically targeted protease, such as a viral protease, is evolved in the presence of a therapeutic protease inhibitor, such as a candidate protease inhibitor or a clinically used protease inhibitor. In some such embodiments, a protease is evolved that exhibits protease activity in the presence of the inhibitor, and thus is resistant to the inhibitor. In some embodiments, the inhibitor-resistant evolved protease comprises at least one mutation observed in inhibitor-resistant proteases in the context of clinical use of the inhibitor. Such evolved proteases are useful for screening compound libraries for inhibitors that are able to inhibit the evolved protease, and can thus serve as therapeutic drugs that inhibit the protease even in its evolved form that is resistant to the initial protease inhibitor.

[00191] Accordingly, some embodiments of this disclosure provide methods for drug screening, comprising providing an evolved protease, contacting it with a candidate protease inhibitor, and determining the activity of the evolved protease in the presence of the candidate inhibitor. The activity of the protease can be determined by any suitable assay, for example, one of the fluorescent or bioluminescent assays described herein. Additional suitable assays for determining protease activity will be apparent to those of skill in the art based on the instant disclosure. In some embodiments, if the activity of the evolved protease is decreased in the presence of the candidate inhibitor as compared to the activity of the protease in the absence of the inhibitor, then the candidate inhibitor is identified as an inhibitor of the evolved protease. In some embodiments, the candidate inhibitor is identified as an inhibitor of the evolved protease if the inhibition in the presence of the candidate inhibitor is at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99%, or at least 99.9% as compared to the activity of the protease in the absence of the candidate inhibitor, or if the protease activity in the presence of the inhibitor is below the detection threshold of the assay used to determine protease activity.

[00192] Some aspects of this disclosure provide methods for using a protease provided herein. In some embodiments, such methods include contacting a protein comprising a protease target cleavage sequence with the protease. In some embodiments, the protein contacted with the protease is a therapeutic target. Exemplary suitable therapeutic targets are provided herein, including, but not limited to, C-C chemokine receptor type 5 (CCR5), Programmed death-ligand 1 (PDL1), Tumor Necrosis Factor alpha (TNFa), Insulin-Degrading Enzyme (IDE), membrane metallo-endopeptidase (MME, Neprilysin), and Interleukin 23 alpha subunit P19 (IL23aP19). Additional suitable proteins that can be targeted with the evolved proteases provided herein will be apparent to those of ordinary skill in the art, and the disclosure is not limited in this respect.

[00193] Sequences of the exemplary proteins listed above and of additional suitable proteins that can be contacted with the evolved proteases provided herein will be apparent to those of ordinary skill in the art. Exemplary sequences include, without limitation:

>gi|7706702|ref|NP_057668.1| interleukin-23 subunit alpha precursor [Homo sapiens]
MLGSRAVMLLLLLPWTAQGRAVPGGSSPAWTQCQQLSQKLCTLAWSAHPLVGHMDLREEGDE
ETTNDVPHIQCGDGCDPQGLRDNSQFCLQRIHQGLIFYEKLLGSDIFTGEPSLLPDSPVGQL
HASLLGLSQLLQPEGHHWETQQIPSLSPSQPWQRLLLRFKILRSLQAFVAVAARVFAHGAAT
LSP (SEQ ID NO: 33)

>gi|154091328|ref|NP_001093638.1| C-C chemokine receptor type 5 [Homo sapiens]
MDYQVSSPIYDINYYTSEPCQKINVKQIAARLLPPLYSLVFIFGFVGNMLVILILINCKRLK
SMTDIYLLNLAISDLFFLLTVPFWAHYAAAQWDFGNTMCQLLTGLYFIGFFSGIFFIILLTI
DRYLAVVHAVFALKARTVTFGVVTSVITWVVAVFASLPGIIFTRSQKEGLHYTCSSHFPYSQ
YQFWKNFQTLKIVILGLVLPLLVMVICYSGILKTLLRCRNEKKRHRAVRLIFTIMIVYFLFW
APYNIVLLLNTFQEFFGLNNCSSSNRLDQAMQVTETLGMTHCCINPIIYAFVGEKFRNYLLV
FFQKHIAKRFCKCCSIFQQEAPERASSVYTRSTGEQEISVGL (SEQ ID NO: 34)

>gi|930425329|ref|NP_001300958.1| programmed cell death 1 ligand 1 isoform c precursor [Homo sapiens]
MRIFAVFIFMTYWHLLNAFTVTVPKDLYVVEYGSNMTIECKFPVEKQLDLAALIVYWEMEDK
NIIQFVHGEEDLKVQHSSYRQRARLLKDQLSLGNAALQITDVKLQDAGVYRCMISYGGADYK
RITVKVNAPYNKINQRILVVDPVTSEHELTCQAEGYPKAEVIWTSSDHQVLSGKTTTTNSKR
EEKLFNVTSTLRINTTTNEIFYCTFRRLDPEENHTAELVIPGNILNVSIKICLTLSPST
(SEQ ID NO: 35)

>gi|25952111|ref|NP_000585.2| tumor necrosis factor [Homo sapiens]
MSTESMIRDVELAEEALPKKTGGPQGSRRCLFLSLFSFLIVAGATTLFCLLHFGVIGPQREE
FPRDLSLISPLAQAVRSSSRTPSDKPVAHVVANPQAEGQLQWLNRRANALLANGVELRDNQL
VVPSEGLYLIYSQVLFKGQGCPSTHVLLTHTISRIAVSYQTKVNLLSAIKSPCQRETPEGAE
AKPWYEPIYLGGVFQLEKGDRLSAEINRPDYLDFAESGQVYFGIIAL (SEQ ID NO: 36)

>gi|155969707|ref|NP_004960.2| insulin-degrading enzyme isoform 1 [Homo sapiens]
MRYRLAWLLHPALPSTFRSVLGARLPPPERLCGFQKKTYSKMNNPAIKRIGNHITKSPEDKR
EYRGLELANGIKVLLISDPTTDKSSAALDVHIGSLSDPPNIAGLSHFCEHMLFLGTKKYPKE

NEYSQFLSEHAGSSNAFTSGEHTNYYFDVSHEHLEGALDRFAQFFLCPLFDESCKDREVNAV
DSEHEKNVMNDAWRLFQLEKATGNPKHPFSKFGTGNKYTLETRPNQEGIDVRQELLKFHSAY
YSSNLMAVCVLGRESLDDLTNLVVKLFSEVENKNVPLPEFPEHPFQEEHLKQLYKIVPIKDI
RNLYVTFPIPDLQKYYKSNPGHYLGHLIGHEGPGSLLSELKSKGWVNTLVGGQKEGARGFMF
FIINVDLTEEGLLHVEDIILHMFQYIQKLRAEGPQEWVFQECKDLNAVAFRFKDKERPRGYT
SKIAGILHYYPLEEVLTAEYLLEEFRPDLIEMVLDKLRPENVRVAIVSKSFEGKTDRTEEWY
GTQYKQEAIPDEVIKKWQNADLNGKFKLPTKNEFIPTNFEILPLEKEATPYPALIKDTAMSK
LWFKQDDKFFLPKACLNFEFFSPFAYVDPLHCNMAYLYLELLKDSLNEYAYAAELAGLSYDL
QNTIYGMYLSVKGYNDKQPILLKKIIEKMATFEIDEKRFEIIKEAYMRSLNNFRAEQPHQHA
MYYLRLLMTEVAWTKDELKEALDDVTLPRLKAFIPQLLSRLHIEALLHGNITKQAALGIMQM
VEDTLIEHAHTKPLLPSQLVRYREVQLPDRGWFVYQQRNEVHNNCGIEIYYQTDMQSTSENM
FLELFCQIISEPCFNTLRTKEQLGYIVFSGPRRANGIQGLRFIIQSEKPPHYLESRVEAFLI
TMEKSIEDMTEEAFQKHIQALAIRRLDKPKKLSAECAKYWGEIISQQYNFDRDNTEVAYLKT
LTKEDIIKFYKEMLAVDAPRRHKVSVHVLAREMDSCPVVGEFPCQNDINLSQAPALPQPEVI
QNMTEFKRGLPLFPLVKPHINFMAAKL (SEQ ID NO: 37)

>gi|116256333|ref|NP_009220.2| neprilysin [Homo sapiens]
MGKSESQMDITDINTPKPKKKQRWTPLEISLSVLVLLLTIIAVTMIALYATYDDGICKSSDC
IKSAARLIQNMDATTEPCTDFFKYACGGWLKRNVIPETSSRYGNFDILRDELEVVLKDVLQE
PKTEDIVAVQKAKALYRSCINESAIDSRGGEPLLKLLPDIYGWPVATENWEQKYGASWTAEK
AIAQLNSKYGKKVLINLFVGTDDKNSVNHVIHIDQPRLGLPSRDYYECTGIYKEACTAYVDF
MISVARLIRQEERLPIDENQLALEMNKVMELEKEIANATAKPEDRNDPMLLYNKMTLAQIQN
NFSLEINGKPFSWLNFTNEIMSTVNISITNEEDVVVYAPEYLTKLKPILTKYSARDLQNLMS
WRFIMDLVSSLSRTYKESRNAFRKALYGTTSETATWRRCANYVNGNMENAVGRLYVEAAFAG
ESKHVVEDLIAQIREVFIQTLDDLTWMDAETKKRAEEKALAIKERIGYPDDIVSNDNKLNNE
YLELNYKEDEYFENIIQNLKFSQSKQLKKLREKVDKDEWISGAAVVNAFYSSGRNQIVFPAG
ILQPPFFSAQQSNSLNYGGIGMVIGHEITHGFDDNGRNFNKDGDLVDWWTQQSASNFKEQSQ
CMVYQYGNFSWDLAGGQHLNGINTLGENIADNGGLGQAYRAYQNYIKKNGEEKLLPGLDLNH
KQLFFLNFAQVWCGTYRPEYAVNSIKTDVHSPGNFRIIGTLQNSAEFSEAFHCRKNSYMNPE
KKCRVW (SEQ ID NO: 38)

[00194] It will be understood that these sequences are exemplary, that additional C-C chemokine receptor type 5 (CCR5), Programmed death-ligand 1 (PDL1), Tumor Necrosis Factor alpha (TNFa), Insulin-Degrading Enzyme (IDE), membrane metallo-endopeptidase (MME, Neprilysin), and Interleukin 23 alpha subunit P19 (IL23aP19) sequences exist, *e.g.*, sequences in other species and sequences comprising SNPs, and that the sequences provided here are not meant to limit the scope of the disclosure.

[00195] In some embodiments, the methods provided herein comprise contacting the target protein with the protease *in vitro*. In some embodiments, the methods provided herein comprise contacting the target protein with the protease *in vivo*. In some embodiments, the methods provided herein comprise contacting the target protein with the protease in a cell, *e.g.*, by delivering the protease or a nucleic acid molecule encoding the protease to the cell. In some embodiments, the methods provided herein comprise contacting the target protein with the protease in a subject, *e.g.*, by administering the protease to the subject, either locally

or systemically. In some such embodiments, the methods is administered to the subject in an amount effective to result in a measurable decrease in the level of full-length target protein in the subject, or in a measurable increase in the level of a cleavage product generated by the protease upon cleavage of the target protein. In some embodiments, the target protein is associated with a disease or disorder and the administration of the protease results in the amelioration of at least one symptom of the disease or disorder.

Host Cells

[00196] Some aspects of this invention relate to host cells for continuous evolution processes as described herein. In some embodiments, a host cell is provided that comprises at least one viral gene encoding a protein required for the generation of infectious viral particles under the control of a conditional promoter, and a fusion protein comprising a transcriptional activator targeting the conditional promoter and fused to an inhibitor via a linker comprising a protease cleavage site. For example, some embodiments provide host cells for phage-assisted continuous evolution processes, wherein the host cell comprises an accessory plasmid comprising a gene required for the generation of infectious phage particles, for example, M13 gIII, under the control of a conditional promoter, as described herein. In some embodiments, the host cells comprises an expression construct encoding a fusion protein as described herein, *e.g.*, on the same accessory plasmid or on a separate vector. In some embodiments, the host cell further provides any phage functions that are not contained in the selection phage, *e.g.*, in the form of a helper phage. In some embodiments, the host cell provided further comprises an expression construct comprising a gene encoding a mutagenesis-inducing protein, for example, a mutagenesis plasmid as provided herein.

[00197] In some embodiments, modified viral vectors are used in continuous evolution processes as provided herein. In some embodiments, such modified viral vectors lack a gene required for the generation of infectious viral particles. In some such embodiments, a suitable host cell is a cell comprising the gene required for the generation of infectious viral particles, for example, under the control of a constitutive or a conditional promoter (*e.g.,* in the form of an accessory plasmid, as described herein). In some embodiments, the viral vector used lacks a plurality of viral genes. In some such embodiments, a suitable host cell is a cell that comprises a helper construct providing the viral genes required for the generation of infectious viral particles. A cell is not required to actually support the life cycle of a viral vector used in the methods provided herein. For example, a cell comprising a gene required for the generation of infectious viral particles under the control of a conditional promoter may

not support the life cycle of a viral vector that does not comprise a gene of interest able to activate the promoter, but it is still a suitable host cell for such a viral vector.

[00198] In some embodiments, the host cell is a prokaryotic cell, for example, a bacterial cell. In some embodiments, the host cell is an *E. coli* cell. In some embodiments, the host cell is a eukaryotic cell, for example, a yeast cell, an insect cell, or a mammalian cell. The type of host cell, will, of course, depend on the viral vector employed, and suitable host cell/viral vector combinations will be readily apparent to those of skill in the art.

[00199] In some embodiments, the viral vector is a phage and the host cell is a bacterial cell. In some embodiments, the host cell is an *E. coli* cell. Suitable *E. coli* host strains will be apparent to those of skill in the art, and include, but are not limited to, New England Biolabs (NEB) Turbo, Top10F', DH12S, ER2738, ER2267, and XL1-Blue MRF'. These strain names are art recognized and the genotype of these strains has been well characterized. It should be understood that the above strains are exemplary only and that the invention is not limited in this respect.

[00200] In some PACE embodiments, for example, in embodiments employing an M13 selection phage, the host cells are *E. coli* cells expressing the Fertility factor, also commonly referred to as the F factor, sex factor, or F-plasmid. The F-factor is a bacterial DNA sequence that allows a bacterium to produce a sex pilus necessary for conjugation and is essential for the infection of *E. coli* cells with certain phage, for example, with M13 phage. For example, in some embodiments, the host cells for M13-PACE are of the genotype F′proA$^+$B$^+$ Δ(lacIZY) zzf::Tn10(TetR)/ endA1 recA1 galE15 galK16 nupG rpsL ΔlacIZYA araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) proBA::pir116 λ.

Kits and apparatuses

[00201] Some aspects of this invention provide kits for continuous protease evolution as described herein. In some embodiments, the kit comprises (a) a vector encoding a phage backbone, for example, an M13 phage backbone, and a multiple cloning site for insertion of a nucleic acid sequence encoding a protease. In some embodiments, the vector or a replication product thereof can be packaged into infectious phage particles in the presence of other phage functions by suitable host cells. In some embodiments, the vector or a replication product thereof lacks at least one gene required for the generation of infectious particles.

[00202] In some embodiments, the kit comprises (b) an accessory plasmid comprising a nucleic acid sequence encoding the at least one gene required for the generation of

infectious particles under the control of a conditional promoter that is activated by a transcriptional activator.

[00203] In some embodiments, the kit comprises (c) an expression construct encoding a fusion protein of the transcriptional activator that activates the promoter of (b) fused to an inhibitor of the transcriptional activator via a linker, and a multiple cloning site for insertion of a nucleic acid sequence encoding a protease cleavage site.

[00204] In some embodiments, the kit further comprises a helper phage providing all phage functions except for the at least one gene required for the generation of infectious phage particles provided by the accessory plasmid of (b). In some embodiments, the helper phage or a replication product thereof cannot be packaged into infectious phage particles.

[00205] In some embodiments, the kit comprises suitable host cells. In some embodiments, the host cells are E. coli host cells. In some embodiments, the kit further comprises a mutagenesis plasmid. In some embodiments, the mutagenesis plasmid comprising a gene expression cassette encoding *umuC* (a components of E. coli translesion synthesis polymerase V), *dam* (deoxyadenosine methylase), and/or *seqA* (a hemimethylated-GATC binding domain), or any combination thereof.

[00206] In some embodiments, a PACE apparatus is provided, comprising a lagoon that is connected to a turbidostat comprising a host cell as described herein. In some embodiments, the host cell is an *E. coli* host cell. In some embodiments, the host cell comprises an accessory plasmid as described herein, a helper plasmid as described herein, a mutagenesis plasmid as described herein, and/or an expression construct encoding a fusion protein as described herein, or any combination thereof. In some embodiments, the lagoon further comprises a selection phage as described herein, for example, a selection phage encoding a protease of interest. In some embodiments, the lagoon is connected to a vessel comprising an inducer for a mutagenesis plasmid, for example, arabinose. In some embodiments, the host cells are *E. coli* cells comprising the F' plasmid, for example, cells of the genotype F′proA$^+$B$^+$ Δ(lacIZY) zzf::Tn10(TetR)/ endA1 recA1 galE15 galK16 nupG rpsL ΔlacIZYA araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) proBA::pir116 λ$^-$.

[00207] For example, in some embodiments, a PACE apparatus is provided, comprising a lagoon of about 100 ml, or about 1 l volume, wherein the lagoon is connected to a turbidostat of about 0.5 l, 1 l , or 3 l volume, and to a vessel comprising an inducer for a mutagenesis plasmid, for example, arabinose, wherein the lagoon and the turbidostat comprise a suspension culture of *E. coli* cells at a concentration of about 5 x 10^8 cells/ml. In some embodiments, the flow of cells through the lagoon is regulated to about 3 lagoon

volumes per hour. In some embodiments, cells are removed from the lagoon by continuous pumping, for example, by using a waste needle set at a height of the lagoon vessel that corresponds to a desired volume of fluid (*e.g.*, about 100 ml, in the lagoon. In some embodiments, the host cells are *E. coli* cells comprising the F' plasmid, for example, cells of the genotype F′proA$^+$B$^+$ Δ(lacIZY) zzf::Tn10(TetR)/ endA1 recA1 galE15 galK16 nupG rpsL ΔlacIZYA araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) proBA::pir116 λ$^-$.

[00208] Some of the embodiments, advantages, features, and uses of the technology disclosed herein will be more fully understood from the Examples below. The Examples are intended to illustrate some of the benefits of the present disclosure and to describe particular embodiments, but are not intended to exemplify the full scope of the disclosure and, accordingly, do not limit the scope of the disclosure.

EXAMPLES

EXAMPLE 1: Phage-Assisted Directed Evolution of Proteases

Transducing protease activity into gene expression

[00209] PACE requires that a target activity be linked to changes in the expression of an essential phage gene such as gene III (gIII). To couple the cleavage of a polypeptide substrate to increases in gene expression, we engineered a PA-RNAP that transduces proteolytic activity into changes in gene expression that are sufficiently strong and rapid to support PACE. T7 RNA polymerase (T7 RNAP) is naturally inhibited when bound to T7 lysozyme[31]. We envisioned that T7 lysozyme could be tethered to T7 RNAP through a flexible linker containing a target protease cleavage site. Without wishing to be bound by theory, it is believed that the effective concentration of the tethered T7 lysozyme with respect to T7 RNAP would be sufficiently high that the T7 RNAP subunit would exist predominantly in the T7 lysozyme-bound, RNAP-inactive state. Proteolysis of the target sequence would disfavor the bound T7 RNAP:T7 lysozyme complex, resulting in the liberation of an active T7 RNAP and expression of gIII placed downstream of a T7 promoter (**FIG. 1A**).

[00210] N-terminal fusions to T7 RNAP are known to be well tolerated, and in the crystal structure of T7 RNAP bound to T7 lysozyme, the C-terminus of T7 lysozyme is only 32 Å from the N-terminus of T7 RNAP, separated by a solvent-exposed channel[32] (**FIG. 5**). In light of this structural information, we linked the two proteins through these proximal termini. Since T7 lysozyme activity is toxic to host *E. coli* cells, we characterized

catalytically inactive lysozyme variants and found that the inactive C131S lysozyme mutant retained its ability to inhibit T7 RNAP without impairing host cell viability.

[00211] To identify T7 RNAP–T7 lysozyme linkers that promote complex formation and result in an inactive polymerase subunit yet permit efficient proteolysis, we screened a small set of linkers consisting of Gly, Ser, and Ala ranging in length from three to ten residues flanking each side of a target protease substrate. We designed PA-RNAP constructs containing linker peptide sequences known to be cleaved by tobacco etch virus (TEV) protease, HCV protease, or human *rhinovirus*-14 3C (HRV) protease. We assayed T7 RNAP activity using a luciferase reporter and observed that T7 lysozyme linked to T7 RNAP through at least 28 residues including the target protease substrate resulted in significant inhibition of RNAP activity (**FIG. 1B**). To assay RNAP activation, we coexpressed each PA-RNAP variant from a plasmid (the complementary plasmid or CP, **FIG. 1C** and **FIG. 6)** together with each of the three proteases (expressed from the expression plasmid or EP, **FIG. 7**) in *E. coli* cells that also harbored a plasmid encoding gIII and luciferase under control of the T7 promoter (the accessory plasmid or AP, **FIG. 1C** and **FIG. 8**).

[00212] Expression of a protease that is not known to cleave the target amino acid sequence in a coexpressed PA-RNAP did not result in enhanced gene expression as measured by luciferase activity (**FIG. 1D**). In contrast, expression of a protease that is known to cleave the target sequence within the PA-RNAP resulted in 18- to 49-fold increase in gene expression for all three cognate combinations of protease and substrate. These data indicate that PA-RNAPs are capable of transducing specific proteolytic cleavage activities into large changes in target gene expression.

Linking protease activity to phage propagation

[00213] Next we sought to use PA-RNAPs to link the life cycle of M13 bacteriophage to protease activity. We generated selection phage (SP) in which gIII was replaced by a gene encoding TEV protease, HCV protease, or HRV protease (**FIG. 9**). Without pIII, these phage are unable to propagate on wild-type E. coli cells. We engineered host E. coli cells containing two plasmids: (*i*) an AP that contains gIII and luciferase under the control of the T7 promoter, and (*ii*) a CP that constitutively expresses a PA-RNAP (**FIG. 1C**). To be sure that the PA-RNAP selection scheme work as intended we analyzed the cleavage of the sensor by Western blot. We observed the loss of the Lysozyme-RNAP fusion and the formation of a new protein that corresponds to the size of T7 RNAP exclusively in the presence of protease phage that recognizes the host encoded PA-RNAP (**FIG. 11**). To assay whether the host cells

could support phage propagation in a protease-dependent manner, we performed activity-dependent plaque assays. We observed that plaque formation, a consequence of phage replication in solid media, only occurred with phage encoding a protease that can cleave the PA-RNAP within the host cells. Phage with mismatched protease/PA-RNAP combinations did not form plaques, indicating that phage encoding non-cognate proteases do not replicate, or replicate at a significantly reduced rate. These observations together establish that the PA-RNAP system is capable of transducing protease activity of a phage-encoded protease into phage production.

[00214] We next tested if the PA-RNAP-based selection supports the continuous propagation of phage encoding active proteases in the continuous liquid culture format required for PACE (**FIG. 2A).** We maintained three host cell cultures, each harboring a CP expressing a PA-RNAP containing one of the three protease cleavage sites (TEV, HCV, or HRV protease substrates), using chemostats diluted with fresh growth media at a fixed rate[30]. Each of these host cell cultures continuously diluted lagoons seeded with various combinations of phage containing TEV, HCV, or HRV protease. Lagoons seeded with phage encoding cognate proteases that can cleave the PA-RNAP within the host cells robustly propagated (10^8-10^{10} pfu mL^{-1} after 72 hours of continuous dilution at 1.0 lagoon volume per hour), while lagoons seeded with phage encoding proteases that do not match the PA-RNAP of incoming host cells washed out ($< 10^4$ pfu mL^{-1}), demonstrating protease activity-dependent propagation in continuous liquid culture.

[00215] In order to determine if this system can selectively replicate phage carrying protease genes with a desired activity at the expense of phage encoding proteases that are unable to cleave the host-cell PA-RNAP, we performed protease phage enrichment experiments in a PACE format. We seeded a lagoon with a 1,000:1 ratio of TEV SP:HCV SP, then allowed the phage to propagate in the lagoon while being continuously diluted with host cells containing a PA-RNAP with the HCV protease recognition site. We periodically sampled the waste line of the lagoon and amplified by PCR the region of the phage containing the protease genes. The TEV protease and HCV protease genes are readily distinguishable as PCR amplicons of distinct lengths. At the start of the experiment the HCV protease phage were virtually undetectable by PCR amplification of the starting population and gel electrophoresis, while TEV protease dominated the lagoon (**FIG. 2B**). After just 24 h of continuous propagation on host cells containing the HCV PA-RNAP, the TEV protease SPs were undetectable, while the HCV protease SPs were strongly enriched ($\geq$ 100,000-fold enrichment over 24 hours).

[00216] We repeated this experiment with a 1,000-fold excess of HCV protease phage over TEV protease phage using host cells containing the TEV protease PA-RNAP (**FIG. 2C**), and a third time using a 1,000-fold excess of TEV protease phage over HRV phage and host cells containing the HRV protease PA-RNAP (**FIG. 2D**). In all three of the enrichment experiments, continuous propagation rapidly and dramatically enriched phage encoding each cognate protease from a minute fraction of the starting phage mixture, while non-cognate proteases washed out of the lagoon (**FIGs. 2A-D**). Collectively, these results indicate that this protease PACE system successfully links specific protease activity to the phage life cycle in a continuous flow format and can strongly and rapidly enrich phage that encode proteases with the ability to cleave a target polypeptide substrate.

Continuous evolution of resistance to HCV protease inhibitors

[00217] As an initial application of protease PACE, we continuously evolved protease enzymes to rapidly assess the drug resistance susceptibility of small-molecule protease inhibitors. Several HCV protease inhibitors are in late-stage clinical trials or are awaiting FDA approval[33,34]. For some HCV protease inhibitor drug candidates, clinically isolated drug resistance mutations are known[20]. First we tested whether small-molecule HCV protease inhibitors can modulate protease activity in the protease PACE system. We observed that the incubation of host cells with either danoprevir (IC_{50} = ~0.3 nM)[35] or asunaprevir (IC_{50} = ~1.0 nM)[36], two second-generation HCV protease inhibitors, inhibited the cellular gene expression arising from the activity of HCV protease on the HCV PA-RNAP in a dose-dependent manner (**FIGs. 3A-B**). These observations suggest that protease inhibitors can create selection pressure during PACE favoring the evolution of protease mutants that retain their ability to cleave a cognate substrate despite the presence of the drug candidates.

[00218] Based on the relationship between protease inhibitor concentration and gene expression in our system (**FIGs. 3A-B**) and initial trial PACE experiments, we selected 20 μM danoprevir as the final concentration to use in the culture media during attempts to continuously evolve drug-resistant HCV proteases. We inoculated two separate lagoons with HCV protease SP and propagated the phage on host cells containing the HCV protease PA-RNAP in the absence of any inhibitor for 6 h to allow the accumulation of mutations in HCV protease genes. Next, we added 20 μM danoprevir to the media that feeds into the host cell culture, and eventually into each of the two replicate lagoons. As a control, we propagated two replicate lagoons of HCV protease phage on HCV protease PA-RNAP host cells with no added protease inhibitor for the same time period. Throughout all of these experiments, we

induced enhanced mutagenesis of the phage genome by activating an improved mutagenesis plasmid (MP) in the host cells with 0.5% arabinose (see Methods and Materials for a description and **Table 1** below for characterization of the improved MP).

	Glucose	Arabinose	Fold above pJC184
None	1.21 ± 1.25	0.26 ± 0.13	-
pJC184	29.6 ± 32.5	2146 ± 2062	1
pAB086k8	10.4 ± 18.5	10330 ± 13862	4.8

Table 1. Frequency of rifampin resistant colonies using the improved MP.

[00219] Phage populations at 6 and 28 h from replicate lagoons were analyzed by high-throughput DNA sequencing. No mutations were substantially enriched in the control lagoons propagated in the absence of any drug candidate (**FIG. 4C**). In contrast, several mutations rapidly evolved in both replicate lagoons in the presence of danoprevir. Mutations at position D168 were predominant among these mutations. By 28 h, lagoon 1 with danoprevir contained 38.8% D168E, 8.3% D168Y, 2.1% D168A, and 1.1% D168V, while lagoon 2 with danoprevir contained 40.3% D168E and 10.7% D168Y (**FIG. 4C**). Other genetic differences between the SPs of these two replicate populations suggest that cross-contamination did not lead to the observed protease variants in these experiments. These findings reveal that the presence of danoprevir caused the population of continuously evolving proteases to rapidly acquire mutations at D168.

[00220] To assay whether the PACE-evolved mutations confer danoprevir drug resistance in HCV protease, we purified recombinant HCV protease variants containing either of the two most highly enriched mutations, D168E and D168Y. Each of these two mutations increase the IC_{50} of danoprevir by ~30-fold (wild-type HCV protease IC_{50} = 1.3 ± 0.1 nM; HCV protease D168E IC_{50} = 38.9 ± 2.4 nM; HCV protease D168Y IC_{50} = 34.4 ± 2.8 nM; IC_{50} ± standard deviation) (**FIG. 4D**). Importantly, the D168E, D168A, and D168V mutations emerging from protease PACE have been previously identified as common drug-resistance mutations in HCV isolated from patients treated with danoprevir[20,37].

[00221] To validate that protease PACE in the presence of a different HCV protease inhibitor can also result in the rapid evolution of drug-resistance mutations, we repeated PACE of HCV protease in the presence of asunaprevir, an HCV protease inhibitor in phase III clinical trials, instead of danoprevir. We selected 75 μM asunaprevir as the final target

concentration to use in the culture media based on dose-dependent gene expression assays (**FIGs. 3A-B**). In order to allow diversity to emerge in the protease population, we first propagated HCV protease phage for 24 h without any inhibitor. Next, to ensure that the populations had sufficient time to evolve mutations that confer drug resistance, we propagated the populations for 24 h with 10 μM asunaprevir, Finally, the asunaprevir concentration was increased to 75 μM for 27 h in order to enrich those mutations that conferred robust drug resistance. HCV protease phage were also propagated for an identical amount of time without any added drug candidate for comparison. High-throughput DNA sequencing of phage populations at the end of the experiment revealed that mutations evolved at substantial levels in the asunaprevir-treated lagoons but not in the control samples (**FIG. 4C**). In this experimental condition as well, mutations at position D168 were highly enriched. In the case of asunaprevir, however, the only substitution at this position to emerge at substantial levels from protease PACE was D168Y, in contrast with the evolution of both D168E and D168Y during protease PACE with danoprevir.

[00222] *In vitro* assays of HCV proteases containing either mutation provides an explanation underlying the strong apparent preference of D168Y over D168E within asunaprevir-treated lagoons. D168Y increases the IC_{50} of asunaprevir by 30-fold, while D168E only increases the IC_{50} of asunaprevir by ~10-fold (wild-type HCV protease IC_{50} = 6.9 ± 0.6 nM; HCV protease D168E IC_{50} = 53.5 ± 3.4 nM; HCV protease D168Y IC_{50} = 214.8 ± 31.9 nM; IC_{50} ± standard deviation) (**FIG. 4E**). Mutations at position D168 have been previously identified in replicon-based asunaprevir resistance experiments[38] and the specific D168Y mutation has been observed to arise in hepatitis C patients treated with asunaprevir[39]. Data from patients also reveals that D168E, D168V, and D168A mutations favor viral escape from danoprevir therapy[37], further supporting that D168 is an important residue for viral escape. Collectively, these results establish that protease PACE in the presence of protease inhibitor drug candidates can very rapidly (1-3 days) reveal clinically relevant mutants that confer strong resistance to the drug candidates, without requiring extensive laboratory or clinical experiments.

Discussion

[00223] Previous efforts to use laboratory evolution to study HCV protease inhibitor resistance have relied on time- and labor-intensive approaches such as viral replication in mammalian cell culture or conventional protein evolution methods, which typically require months to complete[22]. By comparison, the continuous evolution of proteases can reveal key

resistance mutations in as little as ~1 day of PACE. The speed of PACE and its ability to be multiplexed using many lagoons in parallel, each receiving a different drug candidate, and analyzed by high-throughput DNA sequencing of bar-coded lagoon samples, raises the possibility of screening future early-stage hit or lead compounds for their vulnerability to the evolution of drug resistance, before more resource-intensive optimization of *in vivo* properties or clinical trials take place. Rapid and cost-effective access to drug resistance susceptibility enabled by PACE may enhance the more informed selection of more promising early-stage drug candidates for further development. This technique could also be applied to quickly screen a drug candidate across many distinct genotypic variants of a protease target (such as the six major HCV protease genotypes) to reveal each target variant's potential to evolve mutations that abrogate the effectiveness of the drug candidate. As HCV patient isolates and replicon assays have already demonstrated differing drug resistance profiles among different HCV genotypes,[40] this capability could also be used to rapidly identify patient-specific drug treatments that are more likely to offer long-term therapeutic effects on patients infected with specific HCV strains, even in the absence of previous data relating strain genotypes to drug effectiveness.

[00224] The development of protease PACE also expands the scope of PACE to evolve diverse biochemical activities. Prior PACE studies have only evolved RNA polymerases, which have activities that can be directly linked to changes in gIII expression. This study demonstrates how other types of enzymatic activities with no obvious direct connection to gene expression can nevertheless be evolved using PACE by establishing an indirect, but robust, linkage between the activity of interest and gIII expression.

[00225] The protease PACE system provides a strong foundation for the continuous evolution of proteases with reprogrammed specificities. Previous work on reprogramming the DNA substrate selectivity of T7 RNAP enzymes[25,28-30] demonstrated the ability of PACE to rapidly evolve enzymes that accept substrates very different from the native substrate. These reprogramming experiments relied on a "stepping-stone" strategy in which selection phage are transitioned between a series of intermediate substrates[28], and are enhanced by the recent development of modulated selection stringency and negative selection during PACE[30]. In principle these strategies coupled with protease PACE should also enable the continuous evolution of protease enzymes with the tailor-made ability to selectively cleave proteins implicated in human diseases.

Materials and Methods

[00226] **PA-RNAP gene expression response *in vivo*.** All plasmids were constructed by Gibson Assembly 2x Master Mix (NEB); all PCR products were generated using Q5 Hot Start 2x Master Mix (NEB). *E. coli* strain S1030[30] were transformed by electroporation with three plasmids: (*i*) a complementary plasmid (CP) that constitutively expresses a PA-RNAP with one of the three protease cut sites (**FIG. 6**), (*ii*) an accessory plasmid (AP, **FIG. 8**) that encodes gIII-luciferase (translationally coupled) under control of the T7 promoter, and (*iii*) an arabinose-inducible expression plasmid for one of the three proteases (EP, **FIG. 7**). The HRV protease gene was purchased as IDT gblocks and cloned into the expression vector. The MBP-TEV fusion protein was amplified by PCR from pRK793[41]. The MBP fusion was necessary for expression and solubility. The HCV protease gene was generously provided by the Schiffer lab as a constitutively active construct that includes the NS4a cofactor peptide. Cells were grown in 2xYT media to saturation in the presence of antibiotics and 1 mM glucose, then inoculated into 1 mL fresh media containing 1 mM glucose and antibiotics in a 96 well culture plate. After 4.5 h, 150 μL of the cultures were transferred to a black-wall clear-bottom assay plate and luciferase and OD_{600} measurements were taken using a Tecan Infinite Pro plate. The luminescence data was normalized to cell density by dividing by OD_{600}.

[00227] **Protease activity-dependent plaque assays.** Protease phage were cloned using Gibson assembly and the aforementioned expression plasmids as templates. *E. coli* strain S1030 were transformed by electroporation with an AP and a CP. After the transformed host cells were grown in 2xYT to $OD_{600} \sim 1.0$, 100 μL of cells were added to 50 μL of serial dilutions of protease-encoding phage. After one minute, 800 μL of top agar (7g / L agar in 2xYT) was added, mixed and transferred to quarter-plates containing bottom agar (15g / L agar in 2xYT). After overnight incubation at 37 °C, the plates were examined for plaques, which represent zones of slowed growth and diminished turbidity due to phage propagation..

[00228] **PACE propagations and enrichment experiments.** *E. coli* strain S1030 were transformed by electroporation with an AP, CP (one for each of the three PA-RNAPs), and a mutagenesis plasmid (MP, **FIG. 10, Table 1**) encoding arabinose-inducible expression of a dominant-negative mutator variant of *dnaQ*, wild-type *dam*, and wild-type *seqA*[42-44]. Starter cultures were grown overnight in 2xYT supplemented with antibiotics and 1 mM glucose to prevent induction of mutagenesis prior to the PACE experiment. Host cell culture chemostats containing 80 mL of Davis rich media[30] were inoculated with 2 mL of starter

culture and grown at 37 °C with magnetic stir-bar agitation. At approximately OD_{600} 1.0, fresh Davis rich media was pumped in at 80-100 mL h^{-1}, with a chemostat waste needle set at 80 mL. This fixed dilution rate maintains the chemostat culture in late log phase growth, at which point it can be flowed into lagoons seeded with protease phage (initial titers were ~10^5 pfu mL^{-1}). For these experiments, lagoon waste needles were set to maintain a lagoon volume of 15 mL, and host cell cultures were flowed in at 15-17 mL h^{-1}. Arabinose (10% w/v in water) was added directly to lagoons via syringe pump at 0.7 mL h^{-1} to induce mutagenesis. Test propagations were conducted with cognate protease phage as well as non-cognate protease phage. Enrichment experiment lagoons were seeded with 1,000-fold excess of non-cognate protease phage. Lagoon samples were sterile-filtered at least every 24 h, and titers were assessed by plaque assay. Plaque assays were performed with S1030 carrying pJC175e, a plasmid that supplies gIII under control of the phage-shock promoter[30]. Mock selections were monitored by PCR of the protease gene using filtered samples as template and the primers listed below. The distinct sizes of amplicons containing protease genes enabled evaluation of the relative abundance of cognate and non-cognate protease-encoding phage.
Forward primer (BCD582): TGTTTTAGTGTATTCTTTCGCCTCTTTCGTT (SEQ ID NO: 15)
Reverse primer (BCD578): CCCACAAGAATTGAGTTAAGCCCAATAATAAGAGC (SEQ ID NO: 16)

[00229] Inhibition of PA-RNAP response in host *E. coli* cells. Host cells were prepared by electroporation with an AP and the CP encoding the HCV-site PA-RNAP. We prepared 2xYT media with serial dilutions of inhibitors (danoprevir and asunaprevir, MedChemExpress) from stock solutions made in DMSO, and inoculated with a saturated starter culture of host cells. 150 μL cell cultures in a 96-well assay plate were incubated at 37 °C for 1.5 h to allow uptake of inhibitors, then infected with ~10 μL HCV protease phage (multiplicity of infection ~ 10). After 3 h of incubation at 37 °C, the luminescence of each culture was measured on a Tecan Infinite Pro plate reader and normalized to OD_{600}. In the absence of inhibitor, phage-encoded protease will activate the PA-RNAP leading to robust production of luciferase. Relative dose responses to inhibitors compared to control cells without drug were measured in triplicate.

[00230] Evolution of drug resistance in HCV protease using PACE. Host cells were the same as those in the HCV test propagation and enrichment experiments. Chemostats were established in an identical manner as mentioned previously, but with the volume and flow rate halved (40 mL; 40-50 mL h^{-1}). This adjustment was made to provide enough cell culture

to feed two lagoons while also conserving media that contained small-molecule inhibitors. Lagoons were seeded with HCV protease phage and run in duplicate. After 6 h of propagation without any inhibitor, a filtered lagoon sample was taken, and danoprevir was added directly to the chemostat media at 20 μM with 2.5% DMSO to enhance solubility. A final time point was taken after 22 additional hours, and titers were measured by plaque assay on strain S1030 carrying pJC175e. For the asunaprevir experiment, samples were taken every 12 h. After 24 h of propagation with no inhibitor, asunaprevir was added directly to the chemostat media at 10 μM with 2.5% DMSO. After an additional 24 h, asunaprevir dosage was increased to 75 μM and 5% DMSO. Titers were measured by plaque assay on strain S1030 carrying pJC175e.

[00231] High-throughput sequencing of evolved populations. Strain S1030 carrying pJC175e were grown to saturation and used to inoculate fresh media. Host cells were infected with phage samples from the above PACE experiments and incubated for 5 h at 37 °C. DNA from infected cells was extracted using miniprep kits to yield concentrated template phage DNA (Epoch Life Science). PCR reactions were performed using Q5 Hot Start 2x Master Mix (NEB) with a set of tiled primers. The PCR product from the first reaction was diluted ten-fold and 1 μL served as the template for the second PCR. The second PCR added Illumina adapters as well as barcodes; PCR products were purified from agarose gel (Qiagen) and quantified using the Quant-IT Picogreen assay (Invitrogen). Samples were normalized and pooled together to create a sequencing library at approximately 4 nM. The library was quantified by qPCR (KapaBiosystems) and processed by an Illumina MiSeq using the MiSeq Reagent Kit v3 and the 2x300 paired-end protocol. A single paired-end read of 600 bp is sufficient to cover the entire HCV protease gene. Data was analyzed in MATLAB using a custom script that aligned reads to the wild type gene, and filtered out bases that were below a Q-score threshold of 31 (Illumina Q-scores range from 1-40). As a control, the wild-type stock of HCV protease phage was sequenced and processed to assess the extent of sequencing error and PCR bias. At each locus within the gene, the error rate was calculated as the fraction of residues that was not wild-type. This error rate was added to 1% to determine the variant threshold for each locus within the gene.

[00232] Purification and *in vitro* assays of evolved HCV variants. HCV protease variants were sub-cloned by Gibson assembly out of the phage genome and into the previously mentioned EP. EPs were transformed into NEB BL21 DE3 chemically competent cells. Starter cultures were grown to saturation, and 2 mL was used to inoculate 500 mL LB. At OD_{600} = 0.6, cultures were transferred to 20 °C and induced with 0.5% arabinose for 6 h.

Cells were harvested by centrifugation at 5,000 g for 10 m, and resuspended in lysis/bind buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10% glycerol, 5 mM imidazole). Cells were lysed by sonication for a total of 2 m, and then centrifuged for 20 m at 18,000 g to clarify the lysate. Supernatant was flowed through 0.2 mL His-pur nickel resin spin columns that were equilibrated with binding buffer (Pierce-Thermo). Resin was washed with 4 column volumes of wash buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10% glycerol, 20 mM imidazole). HCV protease was eluted in 4 column volumes of 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10% glycerol, 200 mM imidazole. Samples were further purified by size exclusion chromatography on a SuperDex 75 10/300 GL column (GE Healthcare). Size exclusion was performed in 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 10% glycerol, 1 mM DTT. Protein concentrations were determined by UV_{280} on a Nanodrop machine and calculated using an extinction coefficient of 19,000 cm^{-1} M^{-1} and a molecular weight of 23 kDa.

[00233] *In vitro* assays were performed using the commercial HCV RET Substrate 1 (Anaspec), an internally quenched probe that fluoresces upon proteolytic cleavage, according to the manufacturer's instructions. Protease and inhibitors were incubated in assay buffer at room temperature for 5 m prior to addition of substrate. Fluorescence was measured every 30 s for 20 m by a Tecan Infinite Pro plate reader (excitation / emission = 355 nm / 495 nm). Assays were performed at 30 °C with 40 nM protease, 7.5 μM substrate, and varying concentration of inhibitors in a final volume of 100 μL per well in black-wall clear-bottom assay plate. The assay buffer contained 50 mM Tris HCl pH 8.0, 100 mM NaCl, 20% glycerol, 5 mM DTT. Assays were performed in triplicate, and initial reaction velocities were calculated and normalized to controls without inhibitor. The data was fit to the Hill Equation using Igor Pro with base and max parameters fixed at one and zero respectively. The resulting fits yielded IC_{50} values and standard deviations of the estimate.

[00234] **Western blot of PA-RNAP sensor activation.** *E. coli* cells transformed with an AP and a CP were grown to log phase, then infected with a 10-fold excess of protease-encoding phage. After 4.5 h, the cells were harvested by centrifugation at 5000 *g* for 10 min, and then resuspended in LDS Sample Buffer (Life Technologies). Samples were heated to 95 °C for 5 m and vortexed to shear genomic DNA. 4 μL of each sample was loaded onto a protein gel electrophoresis system (Bolt gel system, Life Technologies). The blot was performed using a PVDF membrane (iBlot 2 system, Life Technologies). The membrane was blocked with 5% BSA TBST then incubated overnight with the primary antibody (5% BSA, TBST, 1:5000 anti-T7 RNAP mouse monoclonal, Novagen #70566). The membrane was washed three times, incubated with the secondary antibody (5% BSA, TBST, 1:5000 donkey

anti-mouse, IR-dye conjugate, LI-COR #926-32212) for 60 min, washed three times, then visualized on a LI-COR Odyssey at 800 nm. As seen in **FIG. 11**, the PA-RNAP sensor is proteolyzed to a smaller band of anticipated molecular weight only in the presence of a cognate protease that can cleave the peptide sequence in each PA-RNAP linker.

[00235] Improved mutagenesis plasmid (MP). The previous generation of the mutagenesis plasmid (MP)[25] carried four genes: *dnaQ926*[43] (a dominant-negative *E. coli* DNA polymerase III proofreading subunit), *umuD'* and *umuC* (the components of *E. coli* translesion synthesis polymerase V) and *recA730* (an activated *recA* mutant). The complex of UmuD'$_2$C/RecA730 forms the *E. coli* mutasome complex, a critical requirement for translesion synthesis across predominantly T-T (6-4) photoproducts and pyrimidine dimers[45,46]. Since neither type of mutation is predicted to occur commonly during current PACE experiments, which do not use UV light or chemical mutagens, the genes encoding the mutasome were removed from the MP. To improve the efficiency of mutagenesis, two additional proteins were included on the P_{BAD} transcript: *dam* (deoxyadenosine methylase) and *seqA* (a hemimethylated-GATC binding domain), both of which are known mutators when overexpressed in E. coli[42,44]. The combination of these three genes yielded higher mutagenesis rates in the presence of arabinose during PACE, and resulted in 5-fold higher mutagenesis of *E. coli* chromosomal DNA as assessed by a rifampin resistance assay (see Table 1).

[00236] Rifampin resistance assay. MG1655 Δ*recA E. coli*[47] (CGSC#: 12492) cells were transformed with the appropriate MPs and plated on 2xYT/agar plates supplemented with 40 μg/mL chloramphenicol and 100 mM glucose to ensure that no induction occurs prior to the assay. After overnight growth, single colonies were picked into liquid Davis Rich Media[48] supplemented with 40 μg/mL chloramphenicol and grown for 12-16 hours with vigorous shaking at 37 °C. Cultures were then diluted 1,000-fold in Davis Rich Media and grown until they reached OD_{600} = 0.5-0.7, at which point the were split into two equal volumes, supplemented with either 100 mM glucose or 100 mM arabinose, and allowed to grow for an additional 24 hours. Saturated cultures were serially diluted and plated on 2xYT/agar supplemented with 100 mM glucose with or without 100 μg/mL rifampin. After overnight growth, colonies were counted from both plates, and the frequency of resistant mutants was calculated. This measurement been widely used in the literature as a metric of mutagenesis[49].

[00237] Analysis of high-throughput sequencing data of evolved populations of HCV protease. FASTQ files were automatically generated by the Illumina MiSeq. These

files were already binned by sample barcodes and ready for transfer to a desktop computer for processing via a custom Matlab script. Each read was aligned to the wild-type HCV protease gene in the expected orientation using the Smith-Waterman algorithm. Base calls with Q-scores below a threshold of 31 were converted to ambiguous bases, and the resulting ambiguous codons were turned into a series of three dashes for computationally efficient translation. Ambiguous codons were translated into X's, which were ignored when tabulating allele counts in a matrix. The script automatically cycled through each FASTQ file and saved the resulting allele count matrix in a separate subdirectory. At this stage, matrices for paired-end reads were added together and normalized to yield allele frequencies for each sample.

[00238] We relied on a wild-type control sample to assess PCR and sequencing bias. For this sample, we calculated the frequency of alleles that were not wild-type at each locus to yield the locus-specific error rate. We added 0.01(1%) to the locus-specific error rate to yield our variant call threshold. The allele frequency matrix for each sample was scanned for mutant alleles above the variant call threshold.

[00239] **Matlab Scripts**

Alignment and Tallying Script

```
filename2='ntdHCV.txt';
filename3='aaHCV.txt';
for d=1:48
    orient=orientations{d};
    seqsFile=filenames{d};
    [header,seqs,qscore] = fastqread(seqsFile);
    seqsLength = length(seqs);
    seqsFile = strrep(seqsFile,'.fastq','');
    if exist(seqsFile,'dir');
        error('Directory already exists. Please rename or move it before
moving on.');
    end
    mkdir(seqsFile);
    wtID = -1;
    while wtID < 0
        wtID = fopen(filename2);
    end
    wt = fscanf(wtID,'%s');
    fclose(wtID);
    wtLength = length(wt);
    aaID = -1;
    while aaID < 0
        aaID = fopen(filename3);
    end
    aa = fscanf(aaID,'%s');
    fclose(aaID);
    aaLength = length(aa);
    sBLength = length(seqs);
    nSkips = 0;
    ALN=repmat(' ',[sBLength wtLength]);
    ALNaa=repmat(' ',[sBLength aaLength]);
    for i = 1:sBLength
        if orient=='F'
            [score,alignment,start] = swalign(seqs{i},wt,'Alphabet','NT');
        elseif orient=='R'
             reverse = seqrcomplement(seqs{i});
```

```
            [score,alignment,start] = swalign(reverse,wt,'Alphabet','NT');
            qscore{i}=fliplr(qscore{i});
        end
        len = length(alignment(3,:));
        skip = 0;
        for j = 1:len
            if (alignment(3,j) == '-' || alignment(1,j) == '-')
                skip = 1;
                nSkips = nSkips + 1;
                break;
            end
            if isletter(qscore{i}(start(1)+j-1))
            else
                alignment(1,j) = 'N';
            end
         end
        if skip == 0
            ALN(i,start(2):(start(2)+length(alignment)-1))=alignment(1,:);
            if mod(start(2),3)==1
                frame=1;
            end
            if mod(start(2),3)==2;
                frame=3;
            end
            if mod(start(2),3)==0;
                frame=2;
            end
            ntd=alignment(1,frame:length(alignment));
            index=strfind(ntd,'N');
            modulus=mod([index],3);
            for k=1:length(index)
                ntd(index(k))='-';
                if modulus(k)==1
                    ntd(index(k)+1)='-';
                    ntd(index(k)+2)='-';
                elseif modulus(k)==2 && index(k)>1
                    ntd(index(k)+1)='-';
                    ntd(index(k)-1)='-';
                elseif modulus(k)==0 && index(k)>2
                    ntd(index(k)-2)='-';
                    ntd(index(k)-1)='-';
                end
            end
            aaseq=nt2aa(ntd);

ALNaa(i,round(start(2)/3)+1:round(start(2)/3)+length(aaseq))=aaseq;
        end
    end
    TallyAA=zeros(20,aaLength);
    TallyNTD=zeros(4,wtLength);
    parfor i=1:wtLength

TallyNTD(:,i)=[sum(ALN(:,i)=='A'),sum(ALN(:,i)=='T'),sum(ALN(:,i)=='C'),sum
(ALN(:,i)=='G')];
    end
    parfor i=1:aaLength

TallyAA(:,i)=[sum(ALNaa(:,i)=='A'),sum(ALNaa(:,i)=='R'),sum(ALNaa(:,i)=='N'
),sum(ALNaa(:,i)=='D'),sum(ALNaa(:,i)=='C') ...
            sum(ALNaa(:,i)=='Q'), sum(ALNaa(:,i)=='E'),
sum(ALNaa(:,i)=='G'), sum(ALNaa(:,i)=='H'), sum(ALNaa(:,i)=='I') ...
            sum(ALNaa(:,i)=='L'), sum(ALNaa(:,i)=='K'),
sum(ALNaa(:,i)=='M'), sum(ALNaa(:,i)=='F'), sum(ALNaa(:,i)=='P') ...
            sum(ALNaa(:,i)=='S'), sum(ALNaa(:,i)=='T'),
sum(ALNaa(:,i)=='W'), sum(ALNaa(:,i)=='Y'), sum(ALNaa(:,i)=='V')];
    end
    save(strcat(seqsFile, '/TallyAA'), 'TallyAA');
    save(strcat(seqsFile, '/TallyNTD'), 'TallyNTD');
    dlmwrite(strcat(seqsFile, '/TallyAA.txt'), TallyAA, 'newline', 'pc');
```

```
    dlmwrite(strcat(seqsFile, '/TallyNTD.txt'), TallyNTD, 'precision',
'%.3f', 'newline', 'pc');
end
```

Variant calling Script

```
filename2='ntdHCV.txt';
filename3='aaHCV.txt';
    wtID = -1;
    while wtID < 0
        wtID = fopen(filename2);
    end
    wtNTD = fscanf(wtID,'%s');
    fclose(wtID);
    ntdlength = length(wtNTD);
     aaID = -1;
    while aaID < 0
        aaID = fopen(filename3);
    end
    wtAA = fscanf(aaID,'%s');
    fclose(aaID);
    aalength = length(wtAA);
    iwtAA=aa2int(wtAA);
for d=41:44
    seqsFile1=strrep(filenames{2*d-1},'.fastq','');
    seqsFile2=strrep(filenames{2*d},'.fastq','');
AA1=load(strcat('/MATLAB/',seqsFile1,'/TallyAA.mat'));
    NTD1=load(strcat('/MATLAB/', seqsFile1,'/TallyNTD.mat'));
AA2=load(strcat('/MATLAB/',seqsFile2,'/TallyAA.mat'));
    NTD2=load(strcat('/MATLAB/', seqsFile2,'/TallyNTD.mat'));
    AA=AA1.TallyAA+AA2.TallyAA;
    NTD=NTD1.TallyNTD+NTD2.TallyNTD;
    AA = AA*spdiags((1./sum(AA,1))',0,aalength,aalength);
    NTD = NTD*spdiags((1./sum(NTD,1))',0,ntdlength,ntdlength);
    for i=1:189
        error(i)=1-AA(iwtAA(i),i)+.01;
    end
 save(strcat('/MATLAB/',seqsFile1,'/error'),'error')
end
for d=1:47
    seqsFile1=strrep(filenames{2*d-1},'.fastq','');
    seqsFile2=strrep(filenames{2*d},'.fastq','');
AA1=load(strcat('/MATLAB/',seqsFile1,'/TallyAA.mat'));
    NTD1=load(strcat('/MATLAB/', seqsFile1,'/TallyNTD.mat'));
    AA2=load(strcat('/MATLAB/',seqsFile2,'/TallyAA.mat'));
    NTD2=load(strcat('/MATLAB/', seqsFile2,'/TallyNTD.mat'));
    AA=AA1.TallyAA+AA2.TallyAA;
    NTD=NTD1.TallyNTD+NTD2.TallyNTD;
    AA = AA*spdiags((1./sum(AA,1))',0,aalength,aalength);
    NTD = NTD*spdiags((1./sum(NTD,1))',0,ntdlength,ntdlength);
    mutations='';
    frequencies='';
    for i=1:aalength-1
        index=find(AA(:,i)>error(i));
        notwt=index(index~=iwtAA(i));
        residues=int2aa(notwt);
        if ~isempty(residues);
            string=strcat(wtAA(i), num2str(i-7), residues');
            mutations=strcat(mutations,'_',string);
            frequencies=strcat(frequencies,'_', num2str(AA(notwt,i)'));
        end
    end
    save(strcat('/MATLAB/',seqsFile1,'/AA'),'AA')
    save(strcat('/MATLAB/',seqsFile1,'/mutations'),'mutations')
    save(strcat('/MATLAB/',seqsFile1,'/frequencies'),'frequencies')
end
```

[00240] **Sequences**

[00241] Sequence of complementary plasmid (CP) illustrated in Figure 6 (SEQ ID NO: 17):

ACGGATCGCTTCATGTGGCAGGAGAAAAAAGACTGCACCGGTGCGTCAGCAGAATATGTGATACAGGATATATTC
CGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGCGAGCGGAAATGGCTTACGAACGGGGCGG
AGATTTCCTGGAAGATGCCAGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGG
CTCCGCCCCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTATAAAGA
TACCAGGCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATTCCG
CTGTTATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTG
TATGCACGAACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAAAG
ACATGCAAAAGCACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTT
AAGGCTAAACTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAG
CTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGCGGTTTTTTCGTTTTCAGAGCAAGAGATTACGCGCAGACCAA
AACGATCTCAAGAAGATCATCTTATTAAGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGG
TCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTA
TATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTC
GTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTG
CTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGATTCGAGCT
CGCCCGGGGATCGACCAGTTGGTGATTTTGAACTTTTGCTTTGCCACGGAACGGTCTGCGTTGTCGGGAAGATGC
GTGATCTGATCCTTCAACTCAGCAAAAGTTCGATTTATTCAACAAAGCCGCCGTCCCGTCAAGTCAGCGTAATGC
TCTGCCAGTGTTACAACCAATTAACCAATTCTGATTTCCGGTCAAATAAAACGAAAGGCTCAGTCGAAAGACTGG
GCCTTTCGTTTTGTTTATACATAGGCGAGTACTCTGTTATGGTTCCTTCCTCGAAAGGAAAAAAAAAATGGCTCG
TGTACAGTTTAAACAACGTGAATCTACTGACGCAATCTTTGTTCACTGCTCGGCTACCAAGCCAAGTCAGAATGT
TGGTGTCCGTGAGATTCGCCAGTGGCACAAAGAGCAGGGTTGGCTCGATGTGGGATACCACTTTATCATCAAGCG
AGACGGTACTGTGGAGGCAGGACGAGATGAGATGGCTGTAGGCTCTCACGCTAAGGGTTACAACCACAACTCTAT
CGGCGTCTGCCTTGTTGGTGGTATCGACGATAAAGGTAAGTTCGACGCTAACTTTACGCCAGCCCAAATGCAATC
CCTTCGCTCACTGCTTGTCACACTGCTGGCTAAGTACGAAGGCGCTGTGCTTCGCGCCCATCATGAGGTGGCGCC
GAAGGCTTCCCCTTCGTTCGACCTTAAGCGTTGGTGGGAGAAGAACGAACTGGTCACTTCTGACCGTGGTAGCGG
CGGTGGTGCGAGTGGTGGCGCGCTGGAGGTCCTGTTCCAGGGCCCGGGCGGTAGCGCAGGCAGTGGAGCGGGCGG
TAACACGATTAACATCGCTAAGAACGACTTCTCTGACATCGAACTGGCTGCTATCCCGTTCAACACTCTGGCTGA
CCATTACGGTGAGCGTTTAGCTCGCGAACAGTTGGCCCTTGAGCATGAGTCTTACGAGATGGGTGAAGCACGCTT
CCGCAAGATGTTTGAGCGTCAACTTAAAGCTGGTGAGGTTGCGGATAACGCTGCCGCCAAGCCTCTCATCACTAC
CCTACTCCCTAAGATGATTGCACGCATCAACGACTGGTTTGAGGAAGTGAAAGCTAAGCGCGGCAAGCGCCCGAC
AGCCTTCCAGTTCCTGCAAGAAATCAAGCCGGAAGCCGTAGCGTACATCACCATTAAGACCACTCTGGCTTGCCT
AACCAGTGCTGACAATACAACCGTTCAGGCTGTAGCAAGCGCAATCGGTCGGGCCATTGAGGACGAGGCTCGCTT
CGGTCGTATCCGTGACCTTGAAGCTAAGCACTTCAAGAAAAACGTTGAGGAACAACTCAACAAGCGCGTAGGGCA
CGTCTACAAGAAAGCATTTATGCAAGTTGTCGAGGCTGACATGCTCTCTAAGGGTCTACTCGGTGGCGAGGCGTG
GTCTTCGTGGCATAAGGAAGACTCTATTCATGTAGGAGTACGCTGCATCGAGATGCTCATTGAGTCAACCGGAAT
GGTTAGCTTACACCGCCAAAATGCTGGCGTAGTAGGTCAAGACTCTGAGACTATCGAACTCGCACCTGAATACGC
TGAGGCTATCGCAACCCGTGCAGGTGCGCTGGCTGGCATCTCTCCGATGTTCCAACCTTGCGTAGTTCCTCCTAA
GCCGTGGACTGGCATTACTGGTGGTGGCTATTGGGCTAACGGTCGTCGTCCTCTGGCGCTGGTGCGTACTCACAG
TAAGAAAGCACTGATGCGCTACGAAGACGTTTACATGCCTGAGGTGTACAAAGCGATTAACATTGCGCAAAACAC
CGCATGGAAAATCAACAAGAAAGTCCTAGCGGTCGCCAACGTAATCACCAAGTGGAAGCATTGTCCGGTCGAGGA
CATCCCTGCGATTGAGCGTGAAGAACTCCCGATGAAACCGGAAGACATCGACATGAATCCTGAGGCTCTCACCGC
GTGGAAACGTGCTGCCGCTGCTGTGTACCGCAAGGACAAGGCTCGCAAGTCTCGCCGTATCAGCCTTGAGTTCAT
GCTTGAGCAAGCCAATAAGTTTGCTAACCATAAGGCCATCTGGTTCCCTTACAACATGGACTGGCGCGGTCGTGT
TTACGCTGTGTCAATGTTCAACCCGCAAGGTAACGATATGACCAAAGGACTGCTTACGCTGGCGAAAGGTAAACC
AATCGGTAAGGAAGGTTACTACTGGCTGAAAATCCACGGTGCAAACTGTGCGGGTGTCGATAAGGTTCCGTTCCC
TGAGCGCATCAAGTTCATTGAGGAAAACCACGAGAACATCATGGCTTGCGCTAAGTCTCCACTGGAGAACACTTG
GTGGGCTGAGCAAGATTCTCCGTTCTGCTTCCTTGCGTTCTGCTTTGAGTACGCTGGGGTACAGCACCACGGCCT
GAGCTATAACTGCTCCCTTCCGCTGGCGTTTGACGGGTCTTGCTCTGGCATCCAGCACTTCTCCGCGATGCTCCG
AGATGAGGTAGGTGGTCGCGCGGTTAACTTGCTTCCTAGTGAAACCGTTCAGGACATCTACGGGATTGTTGCTAA
GAAAGTCAACGAGATTCTACAAGCAGACGCAATCAATGGGACCGATAACGAAGTAGTTACCGTGACCGATGAGAA
CACTGGTGAAATCTCTGAGAAAGTCAAGCTGGGCACTAAGGCACTGGCTGGTCAATGGCTGGCTTACGGTGTTAC
TCGCAGTGTGACTAAGCGTTCAGTCATGACGCTGGCTTACGGGTCCAAAGAGTTCGGCTTCCGTCAACAAGTGCT
GGAAGATACCATTCAGCCAGCTATTGATTCCGGCAAGGGTCTGATGTTCACTCAGCCGAATCAGGCTGCTGGATA
CATGGCTAAGCTGATTTGGGAATCTGTGAGCGTGACGGTGGTAGCTGCGGTTGAAGCAATGAACTGGCTTAAGTC
TGCTGCTAAGCTGCTGGCTGCTGAGGTCAAAGATAAGAAGACTGGAGAGATTCTTCGCAAGCGTTGCGCTGTGCA
TTGGGTAACTCCTGATGGTTTCCCTGTGTGGCAGGAATACAAGAAGCCTATTCAGACGCGCTTGAACCTGATGTT

CCTCGGTCAGTTCCGCTTACAGCCTACCATTAACACCAACAAAGATAGCGAGATTGATGCACACAAACAGGAGTC
TGGTATCGCTCCTAACTTTGTACACAGCCAAGACGGTAGCCACCTTCGTAAGACTGTAGTGTGGGCACACGAGAA
GTACGGAATCGAATCTTTTGCACTGATTCACGACTCCTTCGGTACCATTCCGGCTGACGCTGCGAACCTGTTCAA
AGCAGTGCGCGAAACTATGGTTGACACATATGAGTCTTGTGATGTACTGGCTGATTTCTACGACCAGTTCGCTGA
CCAGTTGCACGAGTCTCAATTGGACAAAATGCCAGCACTTCCGGCTAAAGGTAACTTGAACCTCCGTGACATCTT
AGAGTCGGACTTCGCGTTCGCGTAATGGAGATTTTCAACATGCTCCCTCAATCGGTTGAATGTCGCCCTTTTGTC
TTTGGCGCTGGTAAACCATATGAATTTTCTATTGATTGTGACAAAATAAACTTATTCCGTGGTGTCTTTGCGTTT
CTTTTATATGTTGCCACCTTTATGTATGTATTTTCTACGTTTGCTAACATACTGCGTAATAAGGAGTCTTAATCA
TGCCAGTTCTTTTGGGTATTCCGTAGAAAAAGGAAGAGTATGAGGGAAGCGGTGATCGCCGAAGTATCGACTCAA
CTATCAGAGGTAGTTGGCGTCATCGAGCGCCATCTCGAACCGACGTTGCTGGCCGTACATTTGTACGGCTCCGCA
GTGGATGGCGGCCTGAAGCCACACAGTGATATTGATTTGCTGGTTACGGTGACCGTAAGGCTTGATGAAACAACG
CGGCGAGCTTTGATCAACGACCTTTTGGAAACTTCGGCTTCCCCTGGAGAGAGCGAGATTCTCCGCGCTGTAGAA
GTCACCATTGTTGTGCACGACGACATCATTCCGTGGCGTTATCCAGCTAAGCGCGAACTGCAATTTGGAGAATGG
CAGCGCAATGACATTCTTGCAGGTATCTTCGAGCCAGCCACGATCGACATTGATCTGGCTATCTTGCTGACAAAA
GCAAGAGAACATAGCGTTGCCTTGGTAGGTCCAGCGGCGGAGGAACTCTTTGATCCGGTTCCTGAACAGGATCTA
TTTGAGGCGCTAAATGAAACCTTAACGCTATGGAACTCGCCGCCCGACTGGGCTGGCGATGAGCGAAATGTAGTG
CTTACGTTGTCCCGCATTTGGTACAGCGCAGTAACCGGCAAAATCGCGCCGAAGGATGTCGCTGCCGACTGGGCA
ATGGAGCGCCTGCCGGCCCAGTATCAGCCCGTCATACTTGAAGCTAGACAGGCTTATCTTGGACAAGAAGAAGAT
CGCTTGGCCTCGCGCGCAGATCAGTTGGAAGAATTTGTCCACTACGTGAAAGGCGAGATCACCAAGGTAGTCGGC
AAA

rrnb terminator 1318..1362; bla promoter 1363..1392; p15A origin 1..861; sd8 RBS 1404..1417; T7 Lysozyme 1418..1867; Flexible Linker 1868..1897; Protease recognition sequence 1898..1921; Flexible Linker 1922..1951; T7 RNAP 1952..4600; aadA 4840..5628.

[00242] Sequence of expression plasmid (EP) illustrated in Figure 7 (SEQ ID NO: 18):

TGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTG
CCGGATCAAGAGCTACCACCTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTT
CTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTG
TTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAG
GCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGA
TACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGC
AGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTT
CGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAAC
GCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCT
GTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCA
GTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATA
TGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACT
GGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCG
CTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGA
GGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCT
CGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTT
TGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGC
TCACGATACGGGTTACTGATGATGAACAAGAGGACATCCGGTCAAATAAAACGAAAGGCTCAGTCGAAAGACTGG
GCCTTTCGTTTTAGACTTAGGGACCCTTTATGACAACTTGACGGCTACATCATTCACTTTTTCTTCACAACCGGC
ACGGAACTCGCTCGGGCTGGCCCCGGTGCATTTTTTAAATACCCGCGAGAAATAGAGTTGATCGTCAAAACCAAC
ATTGCGACCGACGGTGGCGATAGGCATCCGGGTGGTGCTCAAAAGCAGCTTCGCCTGGCTGATACGTTGGTCCTC
GCGCCAGCTTAAGACGCTAATCCCTAACTGCTGGCGGAAAAGATGTGACAGACGCGACGGCGACAAGCAAACATG
CTGTGCGACGCTGGCGATATCAAAATTGCTGTCTGCCAGGTGATCGCTGATGTACTGACAAGCCTCGCGTACCCG
ATTATCCATCGGTGGATGGAGCGACTCGTTAATCGCTTCCATGCGCCGCAGTAACAATTGCTCAAGCAGATTTAT
CGCCAGCAGCTCCGAATAGCGCCCTTCCCCTTGCCCGGCGTTAATGATTTGCCCAAACAGGTCGCTGAAATGCGG
CTGGTGCGCTTCATCCGGGCGAAAGAACCCCGTATTGGCAAATATTGACGGCCAGTTAAGCCATTCATGCCAGTA
GGCGCGCGGACGAAAGTAAACCCACTGGTGATACCATTCGCGAGCCTCCGGATGACGACCGTAGTGATGAATCTC
TCCTGGCGGGAACAGCAAAATATCACCCGGTCGGCAAACAAATTCTCGTCCCTGATTTTTCACCACCCCCTGACC
GCGAATGGTGAGATTGAGAATATAACCTTTCATTCCCAGCGGTCGGTCGATAAAAAAATCGAGATAACCGTTGGC
CTCAATCGGCGTTAAACCCGCCACCAGATGGGCATTAAACGAGTATCCCGGCAGCAGGGGATCATTTTGCGCTTC
AGCCATACTTTTCATACTCCCACCATTCAGAGAAGAAACCAATTGTCCATATTGCATCAGACATTGCCGTCACTG
CGTCTTTTACTGGCTCTTCTCGCTAACCCAACCGGTAACCCCGCTTATTAAAAGCATTCTGTAACAAAGCGGGAC

CAAAGCCATGACAAAAACGCGTAACAAAAGTGTCTATAATCACGGCAGAAAAGTCCACATTGATTATTTGCACGG
CGTCACACTTTGCTATGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTATCGCAACT
CTCTACTGTTTCTCCATACCCGTTTTTTTACCTGCAGGTGCAGTAAGGAGGAAAAAAAAATGCATCATCATCATC
ATCATGGTGAAAACCTGTATTTTCAGAGTCATATGGCTAGCATGAAAAAAAAAGGATCCGTTGTTATCGTCGGCC
GTATCAACCTGTCCGGTGACACCGCTTACGCTCAGCAGACTCGAGGTGAGGAGGGTTGCCAAGAAACCTCCCAGA
CCGGTCGTGACAAAAACCAGGTTGAAGGTGAAGTTCAGATCGTTTCCACCGCTACCCAGACCTTCCTGGCTACCT
CCATCAACGGTGTTCTGTGGACCGTTTACCACGGTGCTGGTACCCGTACCATCGCTTCCCCGAAAGGTCCGGTTA
CCCAGATGTACACCAACGTTGACAAAGACCTGGTTGGTTGGCAGGCTCCGCAGGGTTCCCGTTCCCTGACCCCGT
GCACCTGCGGTTCCTCCGACCTGTACCTGGTTACCCGTCACGCTGACGTTATCCCGGTTCGTCGTCGTGGTGACT
CCCGTGGTTCCCTGCTGTCCCCGCGTCCGATCTCCTACCTGAAAGGTTCCTCCGGTGGTCCGCTGCTGTGCCCGG
CTGGTCACGCTGTTGGTATCTTCAGGGCTGCTGTTTCCACCCGTGGTGTTGCTAAAGCTGTTGACTTCATCCCGG
TTGAATCCCTGGAAACCACCATGCGTTCCCCGTGACTTAATTAACGGCACTCCTCAGCAAATATAATGACCCTCT
TGATAACCCAAGAGGGCATTTTTTAATGCCCATGGCGTTTATTTGCCGACTACCTTGGTGATCTCGCCTTTCACG
TAGTGGACAAATTCTTCCAACTGATCTGCGCGCGAGGCCAAGCGATCTTCTTCTTGTCCAAGATAAGCCTGTCTA
GCTTCAAGTATGACGGGCTGATACTGGGCCGGCAGGCGCTCCATTGCCCAGTCGGCAGCGACATCCTTCGGCGCG
ATTTTGCCGGTTACTGCGCTGTACCAAATGCGGGACAACGTAAGCACTACATTTCGCTCATCGCCAGCCCAGTCG
GGCGGCGAGTTCCATAGCGTTAAGGTTTCATTTAGCGCCTCAAATAGATCCTGTTCAGGAACCGGATCAAAGAGT
TCCTCCGCCGCTGGACCTACCAAGGCAACGCTATGTTCTCTTGCTTTTGTCAGCAAGATAGCCAGATCAATGTCG
ATCGTGGCTGGCTCGAAGATACCTGCAAGAATGTCATTGCGCTGCCATTCTCCAAATTGCAGTTCGCGCTTAGCT
GGATAACGCCACGGAATGATGTCGTCGTGCACAACAATGGTGACTTCTACAGCGCGGAGAATCTCGCTCTCTCCA
GGGGAAGCCGAAGTTTCCAAAAGGTCGTTGATCAAAGCTCGCCGCGTTGTTTCATCAAGCCTTACGGTCACCGTA
ACCAGCAAATCAATATCACTGTGTGGCTTCAGGCCGCCATCCACTGCGGAGCCGTACAAATGTACGGCCAGCAAC
GTCGGTTCGAGATGGCGCTCGATGACGCCAACTACCTCTGATAGTTGAGTCGATACTTCGGCGATCACCGCTTCC
CTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAA
TGTATTTAGAAAAATAGGCCAAATAGGCCGT

rrnB1 terminator 1318..1362; araC complement(1378..2256); pBAD promoter 2464..2579; SD8 RBS 2595..2609; ColE1 rep_origin 1..588; Rop complement(1016..1207); aadA complement(3339..4130); aadA promoter 4131..4215; tetA/orfL terminator complement(3284..3325); 6xHis-TEVcutsite-HCV protease 2610..3260.

[00243] Sequence of accessory plasmid (AP) illustrated in Figure 8 (SEQ ID NO: 19):

TCAGATCCTTCCGTATTTAGCCAGTATGTTCTCTAGTGTGGTTCGTTGTTTTTGCGTGAGCCATGAGAACGAACC
ATTGAGATCATGCTTACTTTGCATGTCACTCAAAAATTTTGCCTCAAAACTGGTGAGCTGAATTTTTGCAGTTAA
AGCATCGTGTAGTGTTTTTCTTAGTCCGTTACGTAGGTAGGAATCTGATGTAATGGTTGTTGGTATTTTGTCACC
ATTCATTTTTATCTGGTTGTTCTCAAGTTCGGTTACGAGATCCATTTGTCTATCTAGTTCAACTTGGAAAATCAA
CGTATCAGTCGGGCGGCCTCGCTTATCAACCACCAATTTCATATTGCTGTAAGTGTTTAAATCTTTACTTATTGG
TTTCAAAACCCATTGGTTAAGCCTTTTAAACTCATGGTAGTTATTTTCAAGCATTAACATGAACTTAAATTCATC
AAGGCTAATCTCTATATTTGCCTTGTGAGTTTTCTTTTGTGTTAGTTCTTTTAATAACCACTCATAAATCCTCAT
AGAGTATTTGTTTTCAAAAGACTTAACATGTTCCAGATTATATTTTATGAATTTTTTTAACTGGAAAAGATAAGG
CAATATCTCTTCACTAAAAACTAATTCTAATTTTTCGCTTGAGAACTTGGCATAGTTTGTCCACTGGAAAATCTC
AAAGCCTTTAACCAAAGGATTCCTGATTTCCACAGTTCTCGTCATCAGCTCTCTGGTTGCTTTAGCTAATACACC
ATAAGCATTTTCCCTACTGATGTTCATCATCTGAGCGTATTGGTTATAAGTGAACGATACCGTCCGTTCTTTCCT
TGTAGGGTTTTCAATCGTGGGGTTGAGTAGTGCCACACAGCATAAAATTAGCTTGGTTTCATGCTCCGTTAAGTC
ATAGCGACTAATCGCTAGTTCATTTGCTTTGAAAACAACTAATTCAGACATACATCTCAATTGGTCTAGGTGATT
TTAATCACTATACCAATTGAGATGGGCTAGTCAATGATAATTACTAGTCCTTTTCCTTTGAGTTGTGGGTATCTG
TAAATTCTGCTAGACCTTTGCTGGAAAACTTGTAAATTCTGCTAGACCCTCTGTAAATTCCGCTAGACCTTTGTG
TGTTTTTTTTGTTTATATTCAAGTGGTTATAATTTATAGAATAAAGAAAGAATAAAAAAAGATAAAAAGAATAGA
TCCCAGCCCTGTGTATAACTCACTACTTTAGTCAGTTCCGCAGTATTACAAAAGGATGTCGCAAACGCTGTTTGC
TCCTCTACAAAACAGACCTTAAAACCCTAAAGGCTTAAGTAGCACCCTCGCAAGCTCGGGCAAATCGCTGAATAT
TCCTTTTGTCTCCGACCATCAGGCACCTGAGTCGCTGTCTTTTTCGTGACATTCAGTTCGCTGCGCTCACGGCTC
TGGCAGTGAATGGGGGTAAATGGCACTACAGGCGCCTTTTATGGATTCATGCAAGGAAACTACCCATAATACAAG
AAAAGCCCGTCACGGGCTTCTCAGGGCGTTTTATGGCGGGTCTGCTATGTGGTGCTATCTGACTTTTTGCTGTTC
AGCAGTTCCTGCCCTCTGATTTTCCAGTCTGACCACTTCGGATTATCCCGTGACAGGTCATTCAGACTGGCTAAT
GCACCCAGTAAGGCAGCGGTATCATCAACAGGCTTACCCGTCTTACTGTCAAGAGGACATCCGGTCAAATAAAAC
GAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTGCTGAGGAGACTTAGGGACCCTACTAATACGACTCACTAT
AGGGAGAAAGAAGGAGCGACATTGCTCCGTGTATTCACTCGTTGGAATGAATACACAGTGCAGTGTTTATTCTGT

TATTTATGCCAAAAATAAAGGCCACTATCAGGCAGCTTTGTTGTTCTGTTTACCAAGTTCAGGAGGTAACTCATA
AGAAAGACCTGCAGGTGCAGTAAAGGAAAAAAAAAATGAAAAAATTATTATTCGCAATTCCTTTAGTTGTTCCTT
TCTATTCTCACTCCGCTGAAACTGTTGAAAGTTGTTTAGCAAAACCCCATACAGAAAATTCATTTACTAACGTCT
GGAAAGACGACAAAACTTTAGATCGTTACGCTAACTATGAGGGCTGTCTGTGGAATGCTACAGGCGTTGTAGTTT
GTACTGGTGACGAAACTCAGTGTTACGGTACATGGGTTCCTATTGGGCTTGCTATCCCTGAAAATGAGGGTGGTG
GCTCTGAGGGTGGCGGTTCTGAGGGTGGCGGTTCTGAGGGTGGCGGTACTAAACCTCCTGAGTACGGTGATACAC
CTATTCCGGGCTATACTTATATCAACCCTCTCGACGGCACTTATCCGCCTGGTACTGAGCAAAACCCCGCTAATC
CTAATCCTTCTCTTGAGGAGTCTCAGCCTCTTAATACTTTCATGTTTCAGAATAATAGGTTCCGAAATAGGCAGG
GGGCATTAACTGTTTATACGGGCACTGTTACTCAAGGCACTGACCCCGTTAAAACTTATTACCAGTACACTCCTG
TATCATCAAAAGCCATGTATGACGCTTACTGGAACGGTAAATTCAGAGACTGCGCTTTCCATTCTGGCTTTAATG
AGGATCCATTCGTTTGTGAATATCAAGGCCAATCGTCTGACCTGCCTCAACCTCCTGTCAATGCTGGCGGCGGCT
CTGGTGGTGGTTCTGGTGGCGGCTCTGAGGGTGGTGGCTCTGAGGGTGGCGGTTCTGAGGGTGGCGGCTCTGAGG
GAGGCGGTTCCGGTGGTGGCTCTGGTTCCGGTGATTTTGATTATGAAAAGATGGCAAACGCTAATAAGGGGGCTA
TGACCGAAAATGCCGATGAAAACGCGCTACAGTCTGACGCTAAAGGCAAACTTGATTCTGTCGCTACTGATTACG
GTGCTGCTATCGATGGTTTCATTGGTGACGTTTCCGGCCTTGCTAATGGTAATGGTGCTACTGGTGATTTTGCTG
GCTCTAATTCCCAAATGGCTCAAGTCGGTGACGGTGATAATTCACCTTTAATGAATAATTTCCGTCAATATTTAC
CTTCCCTCCCTCAATCGGTTGAATGTCGCCCTTTTGTCTTTGGCGCTGGTAAACCTTACGAGTTCAGTATCGACT
GCGATAAGATCAACCTGTTCCGCGGTGTCTTTGCGTTTCTTTTATATGTTGCCACCTTTATGTATGTATTTTCTA
CGTTTGCTAACATACTGCGTAATAAGGAGTCTTAATGAAATTTGGAAACTTTTTGCTTACATACCAACCTCCCCA
ATTTTCCCAAACAGAGGTAATGAAACGTTTGGTTAAATTAGGTCGCATCTCTGAGGAGTGTGGTTTTGATACCGT
ATGGTTACTGGAGCATCATTTCACGGAGTTTGGTTTGCTTGGTAACCCTTATGTCGCTGCTGCATATTTACTTGG
CGCGACTAAAAAATTGAATGTAGGAACTGCCGCTATTGTTCTTCCCACAGCCCATCCAGTACGCCAACTTGAAGA
TGTGAATTTATTGGATCAAATGTCAAAAGGACGATTTCGGTTTGGTATTTGCCGAGGGCTTTACAACAAGGACTT
TCGCGTATTCGGCACAGATATGAATAACAGTCGCGCCTTAGCGGAATGCTGGTACGGGCTGATAAAGAATGGCAT
GACAGAGGGATATATGGAAGCTGATAATGAACATATCAAGTTCCATAAGGTAAAAGTAAACCCCGCGGCGTATAG
CAGAGGTGGCGCACCGGTTTATGTGGTGGCTGAATCAGCTTCGACGACTGAGTGGGCTGCTCAATTTGGCCTACC
GATGATATTAAGTTGGATTATAAATACTAACGAAAAGAAAGCACAACTTGAGCTTTATAATGAAGTGGCTCAAGA
ATATGGGCACGATATTCATAATATCGACCATTGCTTATCATATATAACATCTGTAGATCATGACTCAATTAAAGC
GAAAGAGATTTGCCGGAAATTTCTGGGGCATTGGTATGATTCTTATGTGAATGCTACGACTATTTTTGATGATTC
AGACCAAACAAGAGGTTATGATTTCAATAAAGGGCAGTGGCGTGACTTTGTATTAAAAGGACATAAAGATACTAA
TCGCCGTATTGATTACAGTTACGAAATCAATCCCGTGGGAACGCCGCAGGAATGTATTGACATAATTCAAAAAGA
CATTGATGCTACAGGAATATCAAATATTTGTTGTGGATTTGAAGCTAATGGAACAGTAGACGAAATTATTGCTTC
CATGAAGCTCTTCCAGTCTGATGTCATGCCATTTCTTAAAGAAAAACAACGTTCGCTATTATATTATGGCGGTGG
CGGTAGCGGCGGTGGCGGTAGCGGCGGTGGCGGTAGCGGCGGTGGCGGTAGCAAATTTGGATTGTTCTTCCTTAA
CTTCATCAATTCAACAACTGTTCAAGAACAGAGTATAGTTCGCATGCAGGAAATAACGGAGTATGTTGATAAGTT
GAATTTTGAACAGATTTTAGTGTATGAAAATCATTTTTCAGATAATGGTGTTGTCGGCGCTCCTCTGACTGTTTC
TGGTTTTCTGCTCGGTTTAACAGAGAAAATTAAAATTGGTTCATTAAATCACATCATTACAACTCATCATCCTGT
CCGCATAGCGGAGGAAGCTTGCTTATTGGATCAGTTAAGTGAAGGGAGATTTATTTTAGGGTTTAGTGATTGCGA
AAAAAAAGATGAAATGCATTTTTTTAATCGCCCGGTTGAATATCAACAGCAACTATTTGAAGAGTGTTATGAAAT
CATTAACGATGCTTTAACAACAGGCTATTGTAATCCAGATAACGATTTTTATAGCTTCCCTAAAATATCTGTAAA
TCCCCATGCTTATACGCCAGGCGGACCTCGGAAATATGTAACAGCAACCAGTCATCATATTGTTGAGTGGGCGGC
CAAAAAAGGTATTCCTCTCATCTTTAAGTGGGATGATTCTAATGATGTTAGATATGAATATGCTGAAAGATATAA
AGCCGTTGCGGATAAATATGACGTTGACCTATCAGAGATAGACCATCAGTTAATGATATTAGTTAACTATAACGA
AGATAGTAATAAAGCTAAACAAGAGACGCGTGCATTTATTAGTGATTATGTTCTTGAAATGCACCCTAATGAAAA
TTTCGAAAATAAACTTGAAGAAATAATTGCAGAAAACGCTGTCGGAAATTATACGGAGTGTATAACTGCGGCTAA
GTTGGCAATTGAAAAGTGTGGTGCGAAAAGTGTATTGCTGTCCTTTGAACCAATGAATGATTTGATGAGCCAAAA
AAATGTAATCAATATTGTTGATGATAATATTAAGAAGTACCACACGGAATATACCTAAACTTAATTAACGGCACT
CCTCAGCAAATATAATGACCCTCTTGATAACCCAAGAGGGCATTTTTTAATGCCCATGGCGTTTACCAATGCTTA
ATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATA
ACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCA
GATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATC
CAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATT
GCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGA
GTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTG
GCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTT
TCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCG
TCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGA
AAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCA
TCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCG
ACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATG
AGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCA

CCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTAGGCCAA
ATAGGCCGT

repA complement(1..951); pSC101 rep_origin complement(955..1700); rrnB1 terminator 1716..1760; T7 promoter 1785..1804; T1 terminator 1812..1957; sd8 RBS 1972..1985; gIII 1986..3260; xluxAB 3260..5383; tetA/orfL bidirectional terminator complement(5408..5449); bla AmpR complement(5463..6323); bla promoter complement(6324..6518).

[00244] Sequence of selection phage (SP) illustrated in Figure 9 (SEQ ID NO: 20):

ATGATTGACATGCTAGTTTTACGATTACCGTTCATCGATTCTCTTGTTTGCTCCAGACTCTCAGGCAATGACCTG
ATAGCCTTTGTAGACCTCTCAAAAATAGCTACCCTCTCCGGCATGAATTTATCAGCTAGAACGGTTGAATATCAT
GTTGATGGTGATTTGACTGTCTCCGGCCTTTCTCACCCTTTTGAATCTTTACCTACACATTACTCAGGCATTGCA
TTTAAAATATATGAGGGTTCTAAAAATTTTTATCCTTGCGTTGAAATAAAGGCTTCTCCCGCAAAAGTATTACAG
GGTCATAATGTTTTTGGTACAACCGATTTAGCTTTATGCTCTGAGGCTTTATTGCTTAATTTTGCTAATTCTTTG
CCTTGCCTGTATGATTTATTGGATGTTAACGCTACTACTATTAGTAGAATTGATGCCACCTTTTCAGCTCGCGCC
CCAAATGAAAATATAGCTAAACAGGTTATTGACCATTTGCGAAATGTATCTAATGGTCAAACTAAATCTACTCGT
TCGCAGAATTGGGAATCAACTGTTACATGGAATGAAACTTCCAGACACCGTACTTTAGTTGCATATTTAAAACAT
GTTGAGCTACAGCACCAGATTCAGCAATTAAGCTCTAAGCCATCCGCAAAAATGACCTCTTATCAAAAGGAGCAA
TTAAAGGTACTCTCTAATCCTGACCTGTTGGAGTTTGCTTCCGGGCTGGTTCGCTTTGAAGCTCGAATTAGAACG
CGATATTTGAAGTCTTTCGGGCTTCCTCTTAATCTTTTTGATGCAATCCGCTTTGCTTCTGACTATAATAGTCAG
GGTAAAGACCTGATTTTTGATTTATGGTCATTCTCGTTTTCTGAACTGTTTAAAGCATTTGAGGGGGATTCAATG
AATATTTATGACGATTCCGCAGTATTGGACGCTATCCAGTCTAAACATTTTACTATTACCCCCTCTGGCAAAACT
TCTTTTGCAAAAGCCTCTCGCTATTTTGGTTTTTATCGTCGTCTGGTAAACGAGGGTTATGATAGTGTTGCTCTT
ACTATGCCTCGTAATTCCTTTTGGCGTTATGTATCTGCATTAGTTGAATGTGGTATTCCTAAATCTCAACTGATG
AATCTTTCTACCTGTAATAATGTTGTTCCGTTAGTTCGTTTTATTAACGTAGATTTTTCTTCCCAACGTCCTGAC
TGGTATAATGAGCCAGTTCTTAAAATCGCATAAGGTAATTCACAATGATTAAAGTTGAAATTAAACCATCTCAAG
CCCAATTTACTACTCGTTCTGGTGTTTCTCGTCAGGGCAAGCCTTATTCACTGAATGAGCAGCTTTGTTACGTTG
ATTTGGGTAATGAATATCCGGTTCTTGTCAAGATTACTCTTGATGAAGGTCAGCCAGCCTATGCGCCTGGTCTGT
ACACCGTTCATCTGTCCTCTTTCAAAGTTGGTCAGTTCGGTTCCCTTATGATTGACCGTCTGCGCCTCGTTCCGG
CTAAGTAACATGGAGCAGGTCGCGGATTTCGACACAATTTATCAGGCGATGATACAAATCTCCGTTGTACTTTGT
TTCGCGCTTGGTATAATCGCTGGGGGTCAAAGATGAGTGTTTTAGTGTATTCTTTCGCCTCTTTCGTTTTAGGTT
GGTGCCTTCGTAGTGGCATTACGTATTTTACCCGTTTAATGGAAACTTCCTCATGAAAAAGTCTTTAGTCCTCAA
AGCCTCTGTAGCCGTTGCTACCCTCGTTCCGATGCTGTCTTTCGCTGCTGAGGGTGACGATCCCGCAAAAGCGGC
CTTTAACTCCCTGCAAGCCTCAGCGACCGAATATATCGGTTATGCGTGGGCGATGGTTGTTGTCATTGTCGGCGC
AACTATCGGTATCAAGCTGTTTAAGAAATTCACCTCGAAAGCAAGCTGATAAACCGATACAATTAAAGGCTCCTT
TTGGAGCCTTTTTTTTCGCGCCAATAAGGAGGAAAAAAAAATGGGCAGCAGCCATCATCATCATCATCACAGCAG
CGGCCTGGTGCCGCGCGGCAGCCATATGGCTAGCATGAAAAAAAAAGGATCCGTTGTTATCGTCGGCCGTATCAA
CCTGTCCGGTGACACCGCTTACGCTCAGCAGACTCGAGGTGAGGAGGGTTGCCAAGAAACCTCCCAGACCGGTCG
TGACAAAAACCAGGTTGAAGGTGAAGTTCAGATCGTTTCCACCGCTACCCAGACCTTCCTGGCTACCTCCATCAA
CGGTGTTCTGTGGACCGTTTACCACGGTGCTGGTACCCGTACCATCGCTTCCCCGAAAGGTCCGGTTACCCAGAT
GTACACCAACGTTGACAAAGACCTGGTTGGTTGGCAGGCTCCGCAGGGTTCCCGTTCCCTGACCCCGTGCACCTG
CGGTTCCTCCGACCTGTACCTGGTTACCCGTCACGCTGACGTTATCCCGGTTCGTCGTCGTGGTGACTCCCGTGG
TTCCCTGCTGTCCCCGCGTCCGATCTCCTACCTGAAAGGTTCCTCCGGTGGTCCGCTGCTGTGCCCGGCTGGTCA
CGCTGTTGGTATCTTCAAGGCTGCTGTTTCCACCCGTGGTGTTGCTAAAGCTGTTGACTTCATCCCGGTTGAATC
CCTGGAAACCACCATGCGTTCCCCGTGATGATGATAATAATGGAGATTTTCAACATGGGCTAGCTCAGCCCTAGG
TATTATGCTAGCGTGGTGTCTGCGTAATAAGGAGTCTTAATCATGCCAGTTCTTTTGGGTATTCCGTTATTATTG
CGTTTCCTCGGTTTCCTTCTGGTAACTTTGTTCGGCTATCTGCTTACTTTTCTTAAAAAGGGCTTCGGTAAGATA
GCTATTGCTATTTCATTGTTTCTTGCTCTTATTATTGGGCTTAACTCAATTCTTGTGGGTTATCTCTCTGATATT
AGCGCTCAATTACCCTCTGACTTTGTTCAGGGTGTTCAGTTAATTCTCCCGTCTAATGCGCTTCCCTGTTTTTAT
GTTATTCTCTCTGTAAAGGCTGCTATTTTCATTTTTGACGTTAAACAAAAAATCGTTTCTTATTTGGATTGGGAT
AAATAATATGGCTGTTTATTTTGTAACTGGCAAATTAGGCTCTGGAAAGACGCTCGTTAGCGTTGGTAAGATTCA
GGATAAAATTGTAGCTGGGTGCAAAATAGCAACTAATCTTGATTTAAGGCTTCAAAACCTCCCGCAAGTCGGGAG
GTTCGCTAAAACGCCTCGCGTTCTTAGAATACCGGATAAGCCTTCTATATCTGATTTGCTTGCTATTGGGCGCGG
TAATGATTCCTACGATGAAAATAAAAACGGCTTGCTTGTTCTCGATGAGTGCGGTACTTGGTTTAATACCCGTTC
TTGGAATGATAAGGAAAGACAGCCGATTATTGATTGGTTTCTACATGCTCGTAAATTAGGATGGGATATTATTTT
TCTTGTTCAGGACTTATCTATTGTTGATAAACAGGCGCGTTCTGCATTAGCTGAACATGTTGTTTATTGTCGTCG
TCTGGACAGAATTACTTTACCTTTTGTCGGTACTTTATATTCTCTTATTACTGGCTCGAAAATGCCTCTGCCTAA

ATTACATGTTGGCGTTGTTAAATATGGCGATTCTCAATTAAGCCCTACTGTTGAGCGTTGGCTTTATACTGGTAA
GAATTTGTATAACGCATATGATACTAAACAGGCTTTTTCTAGTAATTATGATTCCGGTGTTTATTCTTATTTAAC
GCCTTATTTATCACACGGTCGGTATTTCAAACCATTAAATTTAGGTCAGAAGATGAAATTAACTAAAATATATTT
GAAAAAGTTTTCTCGCGTTCTTTGTCTTGCGATTGGATTTGCATCAGCATTTACATATAGTTATATAACCCAACC
TAAGCCGGAGGTTAAAAAGGTAGTCTCTCAGACCTATGATTTTGATAAATTCACTATTGACTCTTCTCAGCGTCT
TAATCTAAGCTATCGCTATGTTTTCAAGGATTCTAAGGGAAAATTAATTAATAGCGACGATTTACAGAAGCAAGG
TTATTCACTCACATATATTGATTTATGTACTGTTTCCATTAAAAAAGGTAATTCAAATGAAATTGTTAAATGTAA
TTAATTTTGTTTTCTTGATGTTTGTTTCATCATCTTCTTTTGCTCAGGTAATTGAAATGAATAATTCGCCTCTGC
GCGATTTTGTAACTTGGTATTCAAAGCAATCAGGCGAATCCGTTATTGTTTCTCCCGATGTAAAAGGTACTGTTA
CTGTATATTCATCTGACGTTAAACCTGAAAATCTACGCAATTTCTTTATTTCTGTTTTACGTGCAAGTAATTTTG
ATATGGTTGGTTCTAACCCTTCCATTATTCAGAAGTATAATCCAAACAATCAGGATTATATTGATGAATTGCCAT
CATCTGATAATCAGGAATATGATGATAATTCCGCTCCTTCTGGTGGTTTCTTTGTTCCGCAAAATGATAATGTTA
CTCAAACTTTTAAAATTAATAACGTTCGGGCAAAGGATTTAATACGAGTTGTCGAATTGTTTGTAAAGTCTAATA
CTTCTAAATCCTCAAATGTATTATCTATTGACGGCTCTAATCTATTAGTTGTTAGTGCACCTAAAGATATTTTAG
ATAACCTTCCTCAATTCCTTTCTACTGTTGATTTGCCAACTGACCAGATATTGATTGAGGGTTTGATATTTGAGG
TTCAGCAAGGTGATGCTTTAGATTTTTCATTTGCTGCTGGCTCTCAGCGTGGCACTGTTGCAGGCGGTGTTAATA
CTGACCGCCTCACCTCTGTTTTATCTTCTGCTGGTGGTTCGTTCGGTATTTTTAATGGCGATGTTTTAGGGCTAT
CAGTTCGCGCATTAAAGACTAATAGCCATTCAAAAATATTGTCTGTGCCACGTATTCTTACGCTTTCAGGTCAGA
AGGGTTCTATCTTTGTTGGCCAGAATGTCCCTTTTATTACTGGTCGTGTGACTGGTGAATCTGCCAATGTAAATA
ATCCATTTCAGACGATTGAGCGTCAAAATGTAGGTATTTCCATGAGCGTTTTTCCTGTTGCAATGGCTGGCGGTA
ATATTGTTCTGGATATTACCAGCAAGGCCGATAGTTTGAGTTCTTCTACTCAGGCAAGTGATGTTATTACTAATC
AAAGAAGTACTGCTACAACGGTTAATTTGCGTGATGGACAGACTCTTTTACTCGGTGGCCTCACTGATTATAAAA
ACACTTCTCAGGATTCTGGCGTACCGTTCCTGTCTAAAATCCCTTTAATCGGCCTCCTGTTTAGCTCCCGCTCTG
ATTCTAACGAGGAAAGCACGTTATACGTGCTCGTCAAAGCAACCATAGTACGCGCCCTGTAGCGGCGCATTAAGC
GCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTC
TTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGA
TTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGA
TAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACA
CTCAACCCTATCTCGGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAG
CTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTAAATATTTGCTTATACAATC
TTCCTGTTTTTGGGGCTTTTCTTATTATCAACCGGGGTACAT

gII 1..1233; gX 898..1233; gV promoter 1208..1237; gV 1245..1508; gVII 1510..1611; gIX 1608..1706; gVIII 1703..1924; gIII promoter 1910..1939; SD8 RBS 1973..1990; HCV protease 1994..2659; J23107 promoter 2683..2712; gVI RBS 2727..2742; gVI 2743..3081; gI promoter 2990..3019; gI 3083..4129; gIV promoter 3961..3990; gIV 4107..5387; packing signal 5386..5463; RNA primer for (-) replication complement(5591..5611); complete (+) origin 5656..5796; essential (+) ori 5656..5699; gII promoter 5821..5849.

[00245] Sequence of mutagenesis plasmid (MP) as illustrated in Figure 10 (SEQ ID NO: 21):

CACTCGGTCGCTACGCTCCGGGCGTGAGACTGCGGCGGGCGCTGCGGACACATACAAAGTTACCCACAGATTCCG
TGGATAAGCAGGGGACTAACATGTGAGGCAAAACAGCAGGGCCGCGCCGGTGGCGTTTTTCCATAGGCTCCGCCC
TCCTGCCAGAGTTCACATAAACAGACGCTTTTCCGGTGCATCTGTGGGAGCCGTGAGGCTCAACCATGAATCTGA
CAGTACGGGCGAAACCCGACAGGACTTAAAGATCCCCACCGTTTCCGGCGGGTCGCTCCCTCTTGCGCTCTCCTG
TTCCGACCCTGCCGTTTACCGGATACCTGTTCCGCCTTTCTCCCTTACGGGAAGTGTGGCGCTTTCTCATAGCTC
ACACACTGGTATCTCGGCTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTAAGCAAGAACTCCCCGTTCAGCCC
GACTGCTGCGCCTTATCCGGTAACTGTTCACTTGAGTCCAACCCGGAAAAGCACGGTAAAACGCCACTGGCAGCA
GCCATTGGTAACTGGGAGTTCGCAGAGGATTTGTTTAGCTAAACACGCGGTTGCTCTTGAAGTGTGCGCCAAAGT
CCGGCTACACTGGAAGGACAGATTTGGTTGCTGTGCTCTGCGAAAGCCAGTTACCACGGTTAAGCAGTTCCCCAA
CTGACTTAACCTTCGATCAAACCACCTCCCCAGGTGGTTTTTTCGTTTACAGGGCAAAAGATTACGCGCAGAAAA
AAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACTGAACCGCTCTAGATTTCAGTGCAATTTATCTCTTCAAAT
GTAGCACCTGAAGTCAGCCCAGGAGGAAGAGGACATCCGGTCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGG
CCTTTCGTTTTAGACTTAGGGACCCTTTATGACAACTTGACGGCTACATCATTCACTTTTTCTTCACAACCGGCA
CGGAACTCGCTCGGGCTGGCCCCGGTGCATTTTTTAAATACCCGCGAGAAATAGAGTTGATCGTCAAAACCAACA

TTGCGACCGACGGTGGCGATAGGCATCCGGGTGGTGCTCAAAAGCAGCTTCGCCTGGCTGATACGTTGGTCCTCG
CGCCAGCTTAAGACGCTAATCCCTAACTGCTGGCGGAAAAGATGTGACAGACGCGACGGCGACAAGCAAACATGC
TGTGCGACGCTGGCGATATCAAAATTGCTGTCTGCCAGGTGATCGCTGATGTACTGACAAGCCTCGCGTACCCGA
TTATCCATCGGTGGATGGAGCGACTCGTTAATCGCTTCCATGCGCCGCAGTAACAATTGCTCAAGCAGATTTATC
GCCAGCAGCTCCGAATAGCGCCCTTCCCCTTGCCCGGCGTTAATGATTTGCCCAAACAGGTCGCTGAAATGCGGC
TGGTGCGCTTCATCCGGGCGAAAGAACCCCGTATTGGCAAATATTGACGGCCAGTTAAGCCATTCATGCCAGTAG
GCGCGCGGACGAAAGTAAACCCACTGGTGATACCATTCGCGAGCCTCCGGATGACGACCGTAGTGATGAATCTCT
CCTGGCGGGAACAGCAAAATATCACCCGGTCGGCAAACAAATTCTCGTCCCTGATTTTTCACCACCCCCTGACCG
CGAATGGTGAGATTGAGAATATAACCTTTCATTCCCAGCGGTCGGTCGATAAAAAAATCGAGATAACCGTTGGCC
TCAATCGGCGTTAAACCCGCCACCAGATGGGCATTAAACGAGTATCCCGGCAGCAGGGGATCATTTTGCGCTTCA
GCCATACTTTTCATACTCCCACCATTCAGAGAAGAAACCAATTGTCCATATTGCATCAGACATTGCCGTCACTGC
GTCTTTTACTGGCTCTTCTCGCTAACCCAACCGGTAACCCCGCTTATTAAAAGCATTCTGTAACAAAGCGGGACC
AAAGCCATGACAAAAACGCGTAACAAAAGTGTCTATAATCACGGCAGAAAAGTCCACATTGATTATTTGCACGGC
GTCACACTTTGCTATGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTATCGCAACTC
TCTACTGTTTCTCCATACCCGTTTTTTTGGACGCGTACAACTCAAGTCTGACATAAATGACCGCTATGAGCACTG
CAATTACACGCCAGATCGTTCTCGCTACCGCAACCACCGGTATGAACCAGATTGGTGCGCACTATGAAGGCCACA
AGATCATTGAGATTGGTGCCGTTGAAGTGGTGAACCGTCGCCTGACGGGCAATAACTTCCATGTTTATCTCAAAC
CCGATCGGCTGGTGGATCCGGAAGCCTTTGGCGTACATGGTATTGCCGATGAATTTTTGCTCGATAAGCCCACGT
TTGCCGAAGTAGCCGATGAGTTCATGGACTATATTCGCGGCGCGGAGTTGGTGATCCATAACGCAGCGTTCGATA
TCGGCTTTATGGACTACGAGTTTTCGTTGCTTAAGCGCGATATTCCGAAGACCAATACTTTCTGTAAGGTCACCG
ATAGCCTTGCGGTGGCGAGGAAAATGTTTCCCGGTAAGCGCAACAGCCTCGATGCGTTATGTGCTCGCTACGAAA
TAGATAACAGTAAACGAACGCTGCACGGGGCATTACTCGATGCCCAGATCCTTGCGGAAGTTTATCTGGCGATGA
CCGGTGGTCAAACGTCGATGGCTTTTGCGATGGAAGGAGAGACACAACAGCAACAAGGTGAAGCAACAATTCAGC
GCATTGTACGTCAGGCAAGTAAGTTACGCGTTGTTTTTGCGACAGATGAAGAGATTGCAGCTCATGAAGCCCGTC
TCGATCTGGTGCAGAAGAAAGGCGGAAGTTGCCTCTGGCGAGCATAATTTAATATCAGTAAACCGGACATAACCC
ATGAAGAAAAATCGCGCTTTTTTGAAGTGGGCAGGGGGCAAGTATCCCCTGCTTGATGATATTAAACGGCATTTG
CCCAAGGGCGAATGTCTGGTTGAGCCTTTTGTAGGTGCCGGGTCGGTGTTTCTCAACACCGACTTTTCTCGTTAT
ATCCTTGCCGATATCAATAGCGACCTGATCAGTCTCTATAACATTGTGAAGATGCGTACTGATGAGTACGTACAG
GCCGCACGCGAGCTGTTTGTTCCCGAAACAAATTGCGCCGAGGTTTACTATCAGTTCCGCGAAGAGTTCAACAAA
AGCCAGGATCCGTTCCGTCGGGCGGTACTGTTTTTATATTTGAACCGCTACGGTTACAACGGCCTGTGTCGTTAC
AATCTGCGCGGTGAGTTTAACGTGCCGTTCGGCCGCTACAAAAAACCCTATTTCCCGGAAGCAGAGTTGTATCAC
TTCGCTGAAAAAGCGCAGAATGCCTTTTTCTATTGTGAGTCTTACGCCGATAGCATGGCGCGCGCAGATGATGCA
TCCGTCGTCTATTGCGATCCGCCTTATGCACCGCTGTCTGCGACCGCCAACTTTACGGCGTATCACACAAACAGT
TTTACGCTTGAACAACAAGCGCATCTGGCGGAGATCGCCGAAGGTCTGGTTGAGCGCCATATTCCAGTGCTGATC
TCCAATCACGATACGATGTTAACGCGTGAGTGGTATCAGCGCGCAAAATTGCATGTCGTCAAAGTTCGACGCAGT
ATAAGCAGCAACGGCGGCACACGTAAAAAGGTGGACGAACTGCTGGCTTTGTACAAACCAGGAGTCGTTTCACCC
GCGAAAAAATAATTCAGCTAAGACACTGCACTGGATTAAGATGAAAACGATTGAAGTTGATGATGAACTCTACAG
CTATATTGCCAGCCACACTAAGCATATCGGCGAGAGCGCATCCGACATTTTACGGCGTATGTTGAAATTTTCCGC
CGCATCACAGCCTGCTGCTCCGGTGACGAAAGAGGTTCGCGTTGCGTCACCTGCTATCGTCGAAGCGAAGCCGGT
CAAAACGATTAAAGACAAGGTTCGCGCAATGCGTGAACTTCTGCTTTCGGATGAATACGCAGAGCAAAAGCGAGC
GGTCAATCGCTTTATGCTGCTGTTGTCTACACTATATTCTCTTGACGCCCAGGCGTTTGCCGAAGCAACGGAATC
GTTGCACGGTCGTACACGCGTTTACTTTGCGGCAGATGAACAAACGCTGCTGAAAAATGGTAATCAGACCAAGCC
GAAACATGTGCCAGGCACGCCGTATTGGGTGATCACCAACACCAACACCGGCCGTAAATGCAGCATGATCGAACA
CATCATGCAGTCGATGCAATTCCCGGCGGAATTGATTGAGAAGGTTTGCGGAACTATCTAAACTTAATTAACGGC
ACTCCTCAGCCAAGTCAAAAGCCTCCGACCGGAGGCTTTTGACTACATGCCCATGGCGTTTACGCCCCGCCCTGC
CACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC
TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATAGTGAAAACGGGGGCGAAGAAG
TTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTC
TCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAAC
TGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA
GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAACTCCGGATGAGCATTCATCAGG
CGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATA
TCCAGCTGAACGGTCTGGTTATAGGTACATTGAGTAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCAT
TGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAT
AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCAAGCCAAA
TAGGCCGT

rrnB1 terminator 867..911; cloDF13 rep_origin complement(39..777); P14/tonB terminator complement(4361..4396); Cat cmR CDS complement(4410..5069); cat promoter 5070..5167; araC complement(927..1805); pBAD promoter 2013..2165; dnaQ926

2166..2897; sd5 RBS 2156..2165; Modified mutS RBS 2907..2925; dam (wt) 2926..3762; seqA (wt) 3791..4336; seqA Native RBS 3771..3790.

EXAMPLE 2: Reprogramming Protease Specificity

[00246] Figure 12 shows the directed evolution of a reprogrammed HRV protease. Wild-type HRV protease cleaves the substrate sequence LEVLFQGP (SEQ ID NO: 22), but does not cleave the target substrate sequence LEVLFQYP (SEQ ID NO: 23). A 96 hour PACE experiment was performed starting with the wild-type protease substrate sequence for 24 hours, followed by a 50:50 mixture of wild-type and target substrate sequence for 24 hours and the target sequence only for 48 hours. After 96 hours, clones comprising mutations T143A and T143P were recovered. Activity of mutant clones on wild-type and target substrate were tested.

[00247] Figure 13 shows the directed evolution of a reprogrammed TEV protease. Wild-type TEV protease cleaves the substrate sequence ENLYFQSA (SEQ ID NO: 24), but does not cleave the target substrate sequence ENLYFESA (SEQ ID NO: 25). A 96 hour PACE experiment was performed starting with the wild-type protease substrate sequence for 24 hours, followed by a 50:50 mixture of wild-type and target substrate sequence for 24 hours and the target sequence only for 48 hours. After 96 hours, clones comprising the following mutations were recovered:

TEV Protease Genotypes at 96h					
	E107D		D148A	N177K	
				N177K	
			D148A	N177K	
				N177K	
				N177K	
Q73R		S135F			stop223
		S135F			stop223
		S135F			stop223

The activity of some mutant TEV clones on wild-type and target substrate were tested.

[00248] In order to improve the specificity of the evolved proteases, a negative selection scheme was employed. Two orthogonal PA-RNAPs were used in parallel: a positive selection PA-RNAP driving PIII from a T7 promoter, and a negative selection PA-RNAP driving PIII-neg from a T3 promoter (Figure 14, upper panel). A population of promiscuous

and specific clones was obtained after positive selection on substrate ENLYaQS starting from a TEV NNK library. The clones exhibited mutations at residues 209, 211, 216, and 218:

A			V209F	W211C	V216F	M218L	
B			V209I	W211I	V216V	M218L	T232P
C			V209I	W211I	V216V	M218L	
D			V209I	W211I	V216V	M218L	
E	P8C		V209F	W221C	V216F	M218L	
F			V209I	W211I	V216V	M218L	
G		P88H	V209I	W211I	V216V	M218L	
H			V209V	W211I	V216L	M218I	

The activity of the TEV clones was assessed (Figure 14, middle panel; "uni…"= uninfected).

[00249] After combined positive selection on ENLYaQS substrate and negative selection on ENLYFQS, only specific genotypes remained in the population of mutant clones:

A		V209I	W211I		V216F	M218W
B	M121L	V209I	W211I		V216F	M218L
C		V209I	W211I		V216F	M218L
D	M121L	V209I	W211I		V216F	M218L
E		V209I	W211I	K215E	V216F	M218L
F	M121L	V209I	W211I		V216F	M218L
G		V209I	W211I		V216F	M218L
H		V209I	W211I		V216F	M218L

The activity of the TEV clones was assessed (Figure 14, lower panel).

[00250] A number of exemplary disease-associated protease target substrate sequences were identified. The target sequences were selected based on exhibiting a high level of homology to a native TEV protease recognition sequence, an extracellular, solvent-exposed localization, conformational flexibility, and conservation of the target substrate sequence between human and animal disease models. A list of monoclonal antibody therapeutics and high-profile targets was compiled, using a TEV specificity matrix to rank all heptapeptides within these proteins. The crystal structures of each of the targets was inspected in order to confirm conformational flexibility and solvent exposure. The following list shows the differences of exemplary target proteins as compared to the native TEV target site (see Dougherty WG *et al.*, *Virology*. 171, 356 (1989).

TEV site	E	X	L	Y	F	Q	S
CCR5	H	F	P	Y	S	Q	Y
PDL1	H	S	S	Y	R	Q	R
TNFa	L	G	G	V	F	Q	L
IDE	E	L	L	K	F	H	S
MME	E	D	E	Y	F	E	N

[00251] TEV proteases that cleave a single mutant substrate were evolved. Figure 15 shows a PACE experiment, in which an initial phase of 24 hours on wild-type substrate (EMLYFQSA, SEQ ID NO: 26) was followed by 24 hours on a 50:50 mixture of wild-type and single-mutant substrate (HNLYFQSA, SEQ ID NO: 27). The following TEV genotypes were observed 24 hours after mixing:

TEV Protease Genotypes 24h After Mixing									
L1	A					N176S			
	B								Q226Stop
	C		S135F						
	D								
	E		S135F						
	F								
	G								
	H		S135F						
L2	A	D90G						N185S	
	B					N176T			
	C				N171D		N177Y		
	D				N171D		N177Y		
	F								
	G			D136E		N176D			

The activity of selected observed TEV genotypes was assessed (Figure 15 lower panel).

[00252] In a separate experiment, TEV proteases that cleave a single mutant substrate were evolved directly for 36 hours on the mutant substrate (Figure 16). The following TEV genotypes were observed:

TEV Protease Genotypes 36h Propagation of NNK Library of Residues 171, 176, 178						
L1	B				N176I	Y178F
	C		R159T	N171D	N176T	
	D		R159T		N176T	
L2	A	I138T		N171D	N176T	
	B	I138T		N171D	N176T	
	C	I138T		N171D	N176T	
	D	I138T		N171D	N176T	

The activity of selected TEV genotypes on different substrates was assessed (Figure 16 lower panel).

[00253] TEV proteases that cleave a double-mutant substrate were evolved using a multi-step evolution strategy as described in Figure 17. The activity of various mutant TEV clones on double-mutant substrates was assessed (Figure 17, lower panel). The following mutant TEV clones were observed:

L6	a		D127A	S135F	N176I	V209D	W211I	V216V	M218Y
	b		D127A	S135F	N176I	V209M	W211I	V216V	M218W
	c		D127A	S135F	N176I	V209D	W211I	V216V	M218Y
	d		D127A	S135F	N176I	V209D	W211I	V216V	M218Y
	e	F91V	D127A	S135F	N176I	V209Q	W211I	V216V	M218W
	f		D127A	S135F	N176I	V209D	W211I	V216V	M218Y
	g		D127A	S135F	N176I	V209V	W211M	V216V	M218F
	h		D127A	S135F	N176I	V209D	W211V	V216V	M218W

	a							S135F		D148A	F162S		N176I					A206T			
	b							S135F		D148A			N176I				R203Q				
	c						F116C	S135F		D148A	F162S		N176I					A206T			
	d							S135F		D148A			N176I								
	e							S135F		D148A			N176I								
	f							S135F					N176I				R203Q				
	g							S135F		D148A			N176I								
L1	h							S135F		D148A			N176I								
	a								I138T				N176T								
	b								I138T				N176T								
	c								I138T				N176T								
	d								I138T				N176T								
	e								I138T				N176T								
	f								I138T				N176T								
	g								I138T				N176T			S181T					
L2	h								I138T				N176T								
	a							S135F						N177K							E223stop
	b	K6E				C110R		S135F	I138T					N177K							E223stop
	c							S135F				N171D		N177K							E223stop
	d	K6E						S135F				N171D		N177K							E223stop
	e	K6E						S135F						N177K							E223stop
	f							S135F				N171D		N177K							E223stop
L3	g							S135F				N171D		N177K							E223stop
	a							S135F					N176T								
	b							S135F					N176T						K215R		
	c				S59L			S135F					N176T							V216stop	
	d							S135F					N176T								
	e							S135F					N176T								
	f							S135F					N176T								
	g							S135F					N176T								
L4	h							S135F					N176T								

[00254] TEV proteases that cleave a triple- mutant substrate were evolved using a multi-step evolution strategy as described in Figure 18. The activity of various mutant TEV clones on triple-mutant substrates was assessed (Figure 18, lower panel). The following mutant TEV clones were observed:

	a						D127A		S135F	T146A	D148P			N176I	N177M		V209M	W211I		M218F			
	b			E106G			D127A		S135F	T146A	D148P			N176I	N177R		V209M	W211I		M218F			
	c						D127A		S135F	T146A	D148P			N176I	N177R		V209M	W211I		M218F			
	d						D127A		S135F	T146R	D148C	H167P		N176I	N177G		V209M	W211I		M218F			
	e						D127A		S135F	T146A	D148P			N176I	N177W		V209M	W211I		M218F			
	f						D127A		S135F	T146C	D148P			N176I	N177M		V209M	W211I		M218F			
	g						D127A		S135F	T146C	D148P				N177M	S200G	V209M	W211I		M218F			
L1	h						D127A		S135F	T146C	D148P			N176I	N177M		V209M	W211I		M218F			
	a						D127A		S135F	T146S	D148P		N171D	N176T	N177M		V209M	W211I		M218F			K229E
	b						D127A		S135F	T146C	D148P		N171D	N176T	N177M		V209E	W211L	V216I	M218W	E223stop		
	c						D127A		S135F	T146S	D148P		N171D	N176T	N177M		V209M	W211I		M218F			K229E
	d						D127A		S135F	T146C	D148P		N171D	N176T	N177M		V209E	W211L	V216I	M218W			
	e		V63I				D127A	F132S	S135F	T146S	D148P		N171D	N176T	N177M		V209M	W211I		M218F			K229E
	f						D127A		S135F	T146S	D148P		N171D	N176T	N177M		V209M	W211I		M218F			K229E
	g						D127A		S135F	T146C	D148P		N171D	N176T	N177M		V209E	W211L	V216I	M218W	E223stop		
L2	h						D127A		S135F	T146C	D148P		N171D	N176T	N177M		V209M	W211I		M218F			K229E
	c				E107D		D127A		S135F	T146A	D148A			N176I			V209S	W211I		M218W			
	d				E107D		D127A		S135F		D148A			N176I			V209E	W211L	V216I	M218W		Q226stop	
	e				E107D		D127A		S135F		D148A			N176I			V209S	W211I		M218W			
	f				E107D		D127A		S135F	T146A	D148A			N176I			V209E	W211L	V216I	M218W			
	g	G32R			E107D		D127A		S135F	T146A	D148A			N176I			V209E	W211L	V216I	M218W		Q226stop	
L3	h				E107D		D127A		S135F		D148A			N176I			V209F	W211C		M218L			

[00255] TEV proteases that cleave a target IL23a peptide were evolved using a multi-step evolution strategy as described in Figure 19. The activity of various mutant TEV clones on IL23a substrates was assessed (Figure 19, lower panel). The following mutant TEV clones were observed:

	a														D127A	S135F	T146A	D148P					N176I	N177W	V209M	W211I	M218F			K229E	Q233stop
	b														D127A	S135F	T146A	D148P					N176I	N177W	V209M	W211I	M218F			K229E	
	c		P8L												D127A	S135F	T146A	D148P					N176I	N177W	V209M	W211I	M218F			K229E	
	d														D127A	S135F	T146A	D148P					N176I	N177W	V209M	W211I	M218F			K229E	Q233stop
	e														D127A	S135F	T146A	D148P					N176I	N177W	V209M	W211I	M218F			K229E	Q233stop
	f														D127A	S135F	T146A	D148P					N176I	N177W	V209M	W211I	M218F			K229E	Q233stop
	g														D127A	S135F	T146A	D148P		R159K			N176I	N177R	V209M	W211I	M218F			K229E	
L1	h														D127A	S135F	T146A	D148P					N176I	N177R	V209M	W211I	M218F			K229E	
	a														D127A	S135F	T146S	D148P			F162S	N171D	N176T	N177M	V209M	W211I	M218F			K229E	
	b														D127A	S135F	T146S	D148P			F162S	N171D	N176T	N177M	V209M	W211I	M218F			K229E	
	c														D127A	S135F	T146S	D148P			F162S	N171D	N176T	N177M	V209M	W211I	M218F			K229E	
	d														D127A	S135F	T146S	D148P			F162S	N171D	N176T	N177M	V209M	W211I	M218F			K229E	
	e														D127A	S135F	T146S	D148P				N171D	N176T	N177M	V209M	W211I	M218F			K229E	
	f														D127A	S135F	T146S	D148P			F162S	N171D	N176T	N177M	V209M	W211I	M218F			K229E	
	g														D127A	S135F	T146S	D148P			F162S	N171D	N176T	N177M	V209M	W211I	M218F	Q226stop		K229E	
L2	h														D127A	S135F	T146S	D148P			F162S	N171D	N176T	N177M	V209M	W211I	M218F			K229E	
	a				N12H								E107D		D127A	S135F	T146A	D148A					N176I		V209F	W211C	M218L	Q226stop			
	b												E107D		D127A	S135F	T146A	D148A					N176I		V209F	W211C	M218L		P227S		
	d												E107D		D127A	S135F	T146A	D148A					N176I		V209F	W211C	M218L				
	e								K67N				E107D		D127A	S135F	T146A	D148A	S152N				N176I		V209F	W211C	M218L				
	f												E107D		D127A	S135F	T146A	D148A					N176I		V209F	W211C	M218L				
	g												E107D		D127A	S135F	T146A	D148A					N176I		V209F	W211C	M218L				
L3	h												E107D		D127A	S135F	T146A	D148A					N176I		V209F	W211C	M218L				
	a											E106G			D127A	S135F	T146S	D148P					N176I	N177F	V209M	W211I	M218F			K229E	
	b													T118S	D127A	S135F	T146A	D148P					N176I	N177R	V209M	W211I	M218F			K229E	
	c														D127A	S135F	T146A	D148P					N176I	N177R	V209M	W211I	M218F			K229E	
	d													T118S	D127A	S135F	T146A	D148P					N176I	N177R	V209M	W211I	M218F			K229E	
	e														D127A	S135F	T146A	D148P					N176I	N177R	V209M	W211I	M218F			K229E	
	f														D127A	S135F	T146A	D148P					N176I	N177R	V209M	W211I	M218F			K229E	
	g				N12T										D127A	S135F	T146A	D148P					N176I	N177R	V209M	W211I	M218F			K229E	
L4	h										R80G				D127A	S135F	T146A	D148P					N176I	N177R	V209M	W211I	M218F			K229E	
	a														D127A	S135F	T146S	D148P				N171D	N176T	N177M	V209M	W211I	M218F			K229E	
	b														D127A	S135F	T146S	D148P				N171D	N176T	N177M	V209M	W211I	M218F			K229E	
	d														D127A	S135F	T146S	D148P				N171D	N176T	N177M	V209M	W211I	M218F			K229E	
	e	K6E													D127A	S135F	T146S	D148P				N171D	N176T	N177M	V209M	W211I	M218F			K229E	
	f					T17A									D127A	S135F	T146S	D148P				N171D	N176T	N177M	V209M	W211I	M218F			K229E	
L5	g									N66D					D127A	S135F	T146S	D148P				N171D	N176T	N177M	V209M	W211I	M218F			K229E	
	a											E106G	E107D		D127A	S135F	T146A	D148A					N176I		V209F	W211C	M218L				
	b												E107D		D127A	S135F	T146A	D148A					N176I				M218F	Q226stop			
	c						T30A	R50K					E107D		D127A	S135F	T146A	D148A					N176I		V209F	W211C	M218L				
	d												E107D		D127A	S135F	T146A	D148A					N176I				M218F	Q226stop			
	e												E107D		D127A	S135F	T146A	D148A					N176I				M218F	Q226stop			
	f												E107D		D127A	S135F	T146A	D148A					N176I				M218F	Q226stop			
	g			R9C									E107D		D127A	S135F	T146A	D148A					N176I				M218F	Q226stop			
L6	h											E106G	E107D		D127A	S135F	T146A	D148A					N176I		V209F	W211C	M218L				

[00256] **Cleavage of MBP-GST Test Substrates.** An MBP-[substrate sequence]-GST fusion protein was digested at 30 °C for 3 hours in 25 mM Tris pH 8.0, 50 mM NaCl, 1 mM EDTA, 20% glycerol. Figure 20 shows the cleavage efficiency of wild-type TEV protease on various substrate sequences. Figure 21 shows the cleavage efficiency of two evolved TEV proteases tested on the same substrates under the same conditions. On the left, cleavage results are shown for evolved protease TEV_L2A (D127A, S135F, **T146S**, **D148P**, F162S, *N171D*, *N176T*, **N177M**, V209M*, W211I*, M218F*, K229E) are shown. On the right, cleavage results are shown for evolved protease TEV_L1 (D127A, S135F, **T146A**, **D148P**, *N176I*, **N177R**, V209M*, W211I*, M218F*, K229E). Mutations that improve solubility and expression are underlined. Mutations relating to P1 are in bold, mutations relating to P2 are marked with an asterisk, and mutations relating to P6 are italicized.

[00257] **Evolution of HCV Proteases that Cleave Macaque MAVS.** Figure 22 shows the evolution of HCV protease variants that cleave macaque MAVS. Exemplary genotypes observed after 48 and 96 hours are displayed below:

48h	a						I132L	K136L	S159H
	b				V51N		I132L	K136L	S159C
	c						I132L	K136L	S159H
	d						I132L	K136L	S159H
	e						I132L	K136L	S159H
	f				V51N		I132L	K136L	S159C
	g			V33I	V51N		I132L	K136L	S159H
	h				V51N		I132L	K136L	S159C
96h	a				V51N		I132L	K136L	S159C
	b		G->S ns4		V51N		I132L	K136L	S159C
	c					T76P	I132L	K136L	S159H
	d				V51N		I132L	K136L	S159C
	e				V51N		I132L	K136L	S159C
	f				V51N		I132L	K136L	S159C
	g	V->I ns4			V51N		I132L	K136L	S159C
	h				V51N		I132L	K136L	S159C

[00258] **Evolution of HCV Proteases that Cleave Ferret MAVS.** Figure 23 shows the evolution of HCV protease variants that cleave ferret MAVS. Exemplary genotypes observed at the various stages during the evolution experiment are displayed.

REFERENCES

1 Schilling, O. & Overall, C. M. Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. Nat Biotechnol 26, 685-694, (2008).

2 Walsh, G. Biopharmaceutical benchmarks 2006. Nat Biotechnol 24, 769-776, (2006).

3 Wehr, M. C. *et al.* Monitoring regulated protein-protein interactions using split TEV. Nat Methods 3, 985-993, (2006).

4 Craik, C. S., Page, M. J. & Madison, E. L. Proteases as therapeutics. Biochem J 435, 1-16, (2011).

5 Gray, D. C., Mahrus, S. & Wells, J. A. Activation of specific apoptotic caspases with an engineered small-molecule-activated protease. Cell 142, 637-646, (2010).

6 von Mehren, M., Adams, G. P. & Weiner, L. M. Monoclonal antibody therapy for cancer. Annu Rev Med 54, 343-369, (2003).

7 Waldmann, T. A. Immunotherapy: past, present and future. Nat Med 9, 269-277, (2003).

8 Caravella, J. & Lugovskoy, A. Design of next-generation protein therapeutics. Curr Opin Chem Biol 14, 520-528, (2010).

9 Russell, A. J. & Fersht, A. R. Rational modification of enzyme catalysis by engineering surface charge. Nature 328, 496-500, (1987).

10 Varadarajan, N., Gam, J., Olsen, M. J., Georgiou, G. & Iverson, B. L. Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity. Proc Natl Acad Sci U S A 102, 6855-6860, (2005).

11 Knight, Z. A., Garrison, J. L., Chan, K., King, D. S. & Shokat, K. M. A remodelled protease that cleaves phosphotyrosine substrates. J Am Chem Soc 129, 11672-11673, (2007).

12 Sellamuthu, S. *et al.* Engineering of protease variants exhibiting altered substrate specificity. Biochem Biophys Res Commun 371, 122-126, (2008).

13 Varadarajan, N., Rodriguez, S., Hwang, B. Y., Georgiou, G. & Iverson, B. L. Highly active and selective endopeptidases with programmed substrate specificities. Nat Chem Biol 4, 290-294, (2008).

14 Yi, L. *et al.* Engineering of TEV protease variants by yeast ER sequestration screening (YESS) of combinatorial libraries. Proc Natl Acad Sci U S A 110, 7229-7234, (2013).

15 Abbenante, G. & Fairlie, D. P. Protease inhibitors in the clinic. Med Chem 1, 71-104, (2005).

16 Turk, B. Targeting proteases: successes, failures and future prospects. Nat Rev Drug Discov 5, 785-799, (2006).

17 Rong, L., Dahari, H., Ribeiro, R. M. & Perelson, A. S. Rapid emergence of protease inhibitor resistance in hepatitis C virus. Sci Transl Med 2, 30ra32, (2010).

18 Ridky, T. & Leis, J. Development of drug resistance to HIV-1 protease inhibitors. J Biol Chem 270, 29621-29623, (1995).

19 Lefebvre, E. & Schiffer, C. A. Resilience to resistance of HIV-1 protease inhibitors: profile of darunavir. AIDS Rev 10, 131-142, (2008).

20 Romano, K. P. *et al.* The molecular basis of drug resistance against hepatitis C virus NS3/4A protease inhibitors. PLoS Pathog 8, e1002832, (2012).

21 Shang, L., Lin, K. & Yin, Z. Resistance mutations against HCV protease inhibitors and antiviral drug design. Curr Pharm Des 20, 694-703, (2014).

22 Verbinnen, T. *et al.* Tracking the evolution of multiple in vitro hepatitis C virus replicon variants under protease inhibitor selection pressure by 454 deep sequencing. J Virol 84, 11124-11133, (2010).

23 Billerbeck, E., de Jong, Y., Dorner, M., de la Fuente, C. & Ploss, A. Animal models for hepatitis C. Curr Top Microbiol Immunol 369, 49-86, (2013).

24 Orencia, M. C., Yoon, J. S., Ness, J. E., Stemmer, W. P. & Stevens, R. C. Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. Nat Struct Biol 8, 238-242, (2001).

25 Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. Nature 472, 499-503, (2011).

26 Hedstrom, L., Szilagyi, L. & Rutter, W. J. Converting trypsin to chymotrypsin: the role of surface loops. Science 255, 1249-1253, (1992).

27 Kim, J. L. *et al.* Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. Cell 87, 343-355, (1996).

28 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 U S A, (2013).

29 Leconte, A. M. *et al.* A population-based experimental model for protein evolution: effects of mutation rate and selection stringency on evolutionary outcomes. Biochemistry 52, 1490-1499, (2013).

30 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).

31 Entus, R., Aufderheide, B. & Sauro, H. M. Design and implementation of three incoherent feed-forward motif based biological concentration sensors. Syst Synth Biol 1, 119-128, (2007).

32 Jeruzalmi, D. & Steitz, T. A. Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. EMBO J 17, 4101-4113, (1998).

33 Clark, V. C., Peter, J. A. & Nelson, D. R. New therapeutic strategies in HCV: second-generation protease inhibitors. Liver Int 33 Suppl 1, 80-84, (2013).

34 Manns, M. P. & von Hahn, T. Novel therapies for hepatitis C - one pill fits all? Nat Rev Drug Discov 12, 595-610, (2013).

35 Jiang, Y. *et al.* Discovery of danoprevir (ITMN-191/R7227), a highly selective and potent inhibitor of hepatitis C virus (HCV) NS3/4A protease. J Med Chem 57, 1753-1769, (2014).

36 Scola, P. M. *et al.* The discovery of asunaprevir (BMS-650032), an orally efficacious NS3 protease inhibitor for the treatment of hepatitis C virus infection. J Med Chem 57, 1730-1752, (2014).

37 Lim, S. R. *et al.* Virologic escape during danoprevir (ITMN-191/RG7227) monotherapy is hepatitis C virus subtype dependent and associated with R155K substitution. Antimicrob Agents Chemother 56, 271-279, (2012).

38 McPhee, F. *et al.* Resistance analysis of the hepatitis C virus NS3 protease inhibitor asunaprevir. Antimicrob Agents Chemother 56, 3670-3681, (2012).

39 McPhee, F. *et al.* Resistance analysis of hepatitis C virus genotype 1 prior treatment null responders receiving daclatasvir and asunaprevir. Hepatology 58, 902-911, (2013).

40 Imhof, I. & Simmonds, P. Genotype differences in susceptibility and resistance development of hepatitis C virus to protease inhibitors telaprevir (VX-950) and danoprevir (ITMN-191). Hepatology 53, 1090-1099, (2011).

41 Kapust, R. B. *et al.* Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Protein Eng 14, 993-1000, (2001).

42 Herman, G. E. & Modrich, P. Escherichia coli K-12 clones that overproduce dam methylase are hypermutable. J Bacteriol 145, 644-646, (1981).

43 Fijalkowska, I. J. & Schaaper, R. M. Mutants in the Exo I motif of Escherichia coli dnaQ: defective proofreading and inviability due to error catastrophe. Proc Natl Acad Sci U S A 93, 2856-2861, (1996).

44 Yang, H., Wolff, E., Kim, M., Diep, A. & Miller, J. H. Identification of mutator genes and mutational pathways in Escherichia coli using a multicopy cloning approach. Mol Microbiol 53, 283-295, (2004).

45 Bagg, A., Kenyon, C. J. & Walker, G. C. Inducibility of a gene product required for UV and chemical mutagenesis in Escherichia coli. Proc Natl Acad Sci U S A 78, 5749-5753, (1981).

46 Burckhardt, S. E., Woodgate, R., Scheuermann, R. H. & Echols, H. UmuD mutagenesis protein of Escherichia coli: overproduction, purification, and cleavage by RecA. Proc Natl Acad Sci U S A 85, 1811-1815, (1988).

47 Mahdi, A. A., Buckman, C., Harris, L. & Lloyd, R. G. Rep and PriA helicase activities prevent RecA from provoking unnecessary recombination during replication fork repair. Genes Dev 20, 2135-2147, (2006).

48 Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A. & Liu, D. R. Negative selection and stringency modulation enable phage-assisted continuous evolution (PACE) of enzymes with altered specificity. Accepted, (2013).

49 Cupples, C. G. & Miller, J. H. A set of lacZ mutations in Escherichia coli that allow rapid detection of each of the six base substitutions. Proc Natl Acad Sci U S A 86, 5345-5349, (1989).

[00259] All publications, patents, patent applications, publication, and database entries (*e.g.*, sequence database entries) mentioned herein, *e.g.*, in the Background, Summary, Detailed Description, Examples, and/or References sections, are hereby incorporated by reference in their entirety as if each individual publication, patent, patent application, publication, and database entry was specifically and individually incorporated herein by reference. In case of conflict, the present application, including any definitions herein, will control.

EQUIVALENTS AND SCOPE

[00260] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents of the embodiments described herein. The scope of the present disclosure is not intended to be limited to the above description, but rather is as set forth in the appended claims.

[00261] Articles such as “a,” “an,” and “the” may mean one or more than one unless indicated to the contrary or otherwise evident from the context. Claims or descriptions that

include "or" between two or more members of a group are considered satisfied if one, more than one, or all of the group members are present, unless indicated to the contrary or otherwise evident from the context. The disclosure of a group that includes "or" between two or more group members provides embodiments in which exactly one member of the group is present, embodiments in which more than one members of the group are present, and embodiments in which all of the group members are present. For purposes of brevity those embodiments have not been individually spelled out herein, but it will be understood that each of these embodiments is provided herein and may be specifically claimed or disclaimed.

[00262] It is to be understood that the invention encompasses all variations, combinations, and permutations in which one or more limitation, element, clause, or descriptive term, from one or more of the claims or from one or more relevant portion of the description, is introduced into another claim. For example, a claim that is dependent on another claim can be modified to include one or more of the limitations found in any other claim that is dependent on the same base claim. Furthermore, where the claims recite a composition, it is to be understood that methods of making or using the composition according to any of the methods of making or using disclosed herein or according to methods known in the art, if any, are included, unless otherwise indicated or unless it would be evident to one of ordinary skill in the art that a contradiction or inconsistency would arise.

[00263] Where elements are presented as lists, *e.g.*, in Markush group format, it is to be understood that every possible subgroup of the elements is also disclosed, and that any element or subgroup of elements can be removed from the group. It is also noted that the term "comprising" is intended to be open and permits the inclusion of additional elements or steps. It should be understood that, in general, where an embodiment, product, or method is referred to as comprising particular elements, features, or steps, embodiments, products, or methods that consist, or consist essentially of, such elements, features, or steps, are provided as well. For purposes of brevity those embodiments have not been individually spelled out herein, but it will be understood that each of these embodiments is provided herein and may be specifically claimed or disclaimed.

[00264] Where ranges are given, endpoints are included. Furthermore, it is to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value within the stated ranges in some embodiments, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise. For purposes of brevity, the values in each range have not been individually spelled out herein,

but it will be understood that each of these values is provided herein and may be specifically claimed or disclaimed. It is also to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values expressed as ranges can assume any subrange within the given range, wherein the endpoints of the subrange are expressed to the same degree of accuracy as the tenth of the unit of the lower limit of the range.

[00265] In addition, it is to be understood that any particular embodiment of the present invention may be explicitly excluded from any one or more of the claims. Where ranges are given, any value within the range may explicitly be excluded from any one or more of the claims. Any embodiment, element, feature, application, or aspect of the compositions and/or methods of the invention, can be excluded from any one or more claims. For purposes of brevity, all of the embodiments in which one or more elements, features, purposes, or aspects is excluded are not set forth explicitly herein.

CLAIMS

What is claimed is:

1. A method of protease evolution, comprising

(a) contacting a population of host cells with a population of phage vectors comprising a gene encoding a protease and deficient in at least one gene for the generation of infectious phage particles, wherein

(1) the host cells are amenable to transfer of the vector;

(2) the vector allows for expression of the protease in the host cell, can be replicated by the host cell, and the replicated vector can transfer into a second host cell;

(3) the host cell expresses a gene product encoded by the at least one gene for the generation of infectious phage particles of (a) in response to the activity of the protease, and the level of gene product expression depends on the activity of the protease;

(b) incubating the population of host cells under conditions allowing for mutation of the gene encoding the protease and the transfer of the vector comprising the gene encoding the protease of interest from host cell to host cell, wherein host cells are removed from the host cell population, and the population of host cells is replenished with fresh host cells that do not harbor the vector;

(c) isolating a replicated vector from the host cell population in (b), wherein the replicated vector comprises a mutated version of the gene encoding the protease.

2. The method of claim 1, wherein the host cell expresses a fusion molecule comprising

(i) a transcriptional activator; and

(ii) an inhibitor of the transcriptional activator of (i), wherein the inhibitor is fused to the transcriptional activator of (i) via a linker comprising a protease cleavage site that is cleaved by the protease of (a).

3. The method of claim 2, wherein the at least one gene for the generation of infectious phage particles is expressed under the control of a promoter activated by the transcriptional activator of (i).

4. The method of any one of claims 2-3, wherein the inhibitor of (ii) is fused to the N-terminus of the transcriptional activator of (i).

5. The method of any one of claims 2-4, wherein the inhibitor of (ii) is fused to the C-terminus of the transcriptional activator of (i).

6. The method of any one of claims 2-5, wherein the transcriptional activity of the fusion protein is inhibited as compared to the activity of the transcriptional activator alone.

7. The method of any one of claims 2-6, wherein the transcriptional activity of the fusion protein is less than 50%, less than 25%, less than 20%, less than 10%, less than 5%, less than 4%, less than 3%, less than 1%, less than 0.1%, less than 0.01%, or less than 0.001% the activity of the transcriptional activator alone.

8. The method of any one of claims 2-7, wherein the transcriptional activity of the fusion protein is assessed in a fluorescent or bioluminescent assay.

9. The method of any one of claims 2-8, wherein cleavage of the protease cleavage site results in an activation of the transcriptional activator.

10. The method of any one of claims 2-9, wherein the activity of the transcriptional activator comprised in the fusion protein is increased at least2-fold, at least5-fold, at least10-fold, at least20-fold, at least25-fold, at least 50-fold, at least 75-fold, at least 100-fold, at least 200-foled, at least 250-fold, at least 500-fold, at least 750-fold, at least 1000-fold, at least2000-fold, at least2500-fold, at least 5000-fold, at least 7500-fold, or at least 10000-fold upon cleavage of the protease cleavage site of the linker.

11. The method of any one of claims 1-10, wherein the protease is a therapeutic protease.

12. The method of any one of claims 1-11, wherein the replicated vector isolated in (c) encodes a mutated protease that cleaves the protease cleavage site with higher efficiency than the version of the protease from (a).

13. The method of any one of claims 1-10, wherein the protease is a target of protease inhibitor.

14. The method of claim 13, wherein the protease inhibitor is a therapeutic agent.

15. The method of any one of claims 13-14, wherein step (b) is carried out in the presence of the protease inhibitor.

16. The method of any one of claims 13-15, wherein the protease inhibitor is used at a concentration that does not completely inhibit the activity of the protease.

17. The method of any one of claims 13-15, wherein the protease inhibitor is used at a concentration that results in an inhibition of at least 50%, at least 75%, at least 80%, at least 90%, at least 95%, at least 97.5%, at least 98%, at least 99%, at least 99.5%, or at least 99.9% of the protease activity.

18. The method of any one of claims 13-17, wherein the protease inhibitor is used at an initial concentration that does not completely inhibit the activity of the protease, and wherein the concentration of the inhibitor is subsequently raised to increase the level of protease inhibition.

19. The method of any one of claims 13-18, wherein the replicated vector isolated in (c) encodes a mutated protease that cleaves the protease cleavage site in the presence of the inhibitor.

20. The method of any one of claims 13-19, wherein the replicated vector isolated in (c) encodes a mutated protease that cleaves the protease cleavage site in the presence of the inhibitor with higher efficiency than the version of the protease from (a).

21. The method of any one of claims 2-20, wherein the transcriptional activator is a transcription factor.

22. The method of claim 21, wherein the transcriptional activator is an RNA polymerase.

23. The method of claim 21, wherein the transcriptional activator is a T7 RNA polymerase.

24. The method of any one of claims 1-23, wherein the vector is a viral vector.

25. The method of claim 24, wherein the viral vector is a phage.

26. The method of claim 25, wherein the phage is a filamentous phage.

27. The method of claim 26, wherein the phage is an M13 phage.

28. The method of any one of claims 1-27, wherein the at least one gene for the generation of infectious phage particles comprises a sequence encoding a pIII protein.

29. The method of any one of claims 1-28, wherein the at least one gene for the generation of infectious phage particles comprises a full-length gIII gene.

30. The method of any one of claims 25-29, wherein the host cells comprise all phage genes except for the at least one gene for the generation of infectious phage particles in the form of a helper phage.

31. The method of any one of claims 25-30, wherein the phage genes comprised on the helper phage comprise pI, pII, pIV, pV, pVI, pVII, pVIII, pIX, and/or pX.

32. The method of any one of claims 1-31, wherein the host cells comprise an accessory plasmid.

33. The method of claim 32, wherein the accessory plasmid comprises an expression construct encoding the pIII protein under the control of a promoter that is activated by the transcriptional activator comprised in the fusion protein.

34. The method of any one of claims 25-33, wherein the host cells comprise the accessory plasmid and together, the helper phage and the accessory plasmid comprise all genes required for the generation of an infectious phage.

35. The method of any of claims 1-34, wherein the method further comprises a negative selection for undesired protease activity.

36. The method of claim 35, wherein the host cells comprise an expression construct encoding a dominant-negative pIII protein (pIII-neg).

37. The method of claim 36, wherein expression of the pIII-neg protein is driven by a promoter the activity of which depends on an undesired function of the protease.

38. The method of claim 37, wherein the undesired function is cleavage of an off-target protease cleavage site.

39. The method of claim 37, wherein the undesired function is cleavage of the linker sequence of the fusion protein outside of the protease cleavage site.

40. The method of any one of claims 1-39, wherein the host cells further comprise a mutagenesis plasmid.

41. The method of claim 40, wherein the mutagenesis plasmids comprises a gene expression cassette encoding a mutagenesis-promoting gene product.

42. The method of claim 40 or 41, wherein the mutagenesis plasmid comprises a gene expression cassette encoding *umuC* (a component of E. coli translesion synthesis polymerase V), *dam* (deoxyadenosine methylase), or *seqA* (a hemimethylated-GATC binding domain), or any combination thereof.

43. The method of any one of claims 40-42, wherein the mutagenesis plasmid further comprises a nucleic acid encoding , UmuD′, and/or RecA.

44. The method of any one of claims 40-43, wherein the expression cassette comprises a conditional promoter, the activity of which depends on the presence of an inducer.

45. The method of claim 44, wherein the conditional promoter is an arabinose-inducible promoter and the inducer is arabinose.

46. The method of claim 40-45, further comprising contacting the host cells with an inducer that induces the expression of the mutagenesis-promoting genes in the host cells.

47. The method of any of claims 24-46, wherein step (b) comprises incubating the population of host cells for a time sufficient for at least 10, at least 20, at least 30, at least 40, at least 50, at least 100, at least 200, at least 300, at least 400, at least, 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1250, at least 1500, at least 1750, at least 2000, at least 2500, at least 3000, at least 4000, at least 5000, at least 7500, at least 10000, or more consecutive life cycles of the viral vector or phage.

48. The method of any of claims 1-47, wherein the host cells are E. coli cells.

49. The method of any of claims 1-48, wherein the host cells are incubated in suspension culture.

50. The method of any of claims 1-49, wherein the population of host cells is continuously replenished with fresh host cells that do not comprise the vector.

51. The method of any of claims 1-50, wherein fresh cells are being replenished and cells are being removed from the cell population at a rate resulting in a substantially constant number of cells in the cell population.

52. The method of any of claims 1-51, wherein fresh cells are being replenished and cells are being removed from the cell population at a rate resulting in a substantially constant vector population.

53. The method of any of claims 1-52, wherein fresh cells are being replenished and cells are being removed from the cell population at a rate resulting in a substantially constant vector, viral, or phage load.

54. The method of any of claims 1-53, wherein the rate of fresh cell replenishment and/or the rate of cell removal is adjusted based on quantifying the cells in the cell population.

55. The method of any of claims 1-54, wherein the rate of fresh cell replenishment and/or the rate of cell removal is adjusted based on quantifying the frequency of host cells harboring the vector and/or of host cells not harboring the vector in the cell population.

56. The method of any of claims 54-55, wherein the quantifying is by measuring the turbidity of the host cell culture, measuring the host cell density, measuring the wet weight of host cells per culture volume, or by measuring light extinction of the host cell culture.

57. The method of any of claims 1-56, wherein the host cells are exposed to a mutagen.

58. The method of claim 57, wherein the mutagen is ionizing radiation, ultraviolet radiation, base analogs, deaminating agents (*e.g.*, nitrous acid), intercalating agents (*e.g.*, ethidium bromide), alkylating agents (*e.g.*, ethylnitrosourea), transposons, bromine, azide salts, psoralen, benzene,3- Chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone (MX) (CAS no. 77439-76-0), O,O-dimethyl-S-(phthalimidomethyl)phosphorodithioate (phos-met) (CAS no. 732-11- 6), formaldehyde (CAS no. 50-00-0), 2-(2-furyl)-3-(5-nitro-2-furyl)acrylamide (AF-2) (CAS no. 3688-53-7), glyoxal (CAS no. 107-22-2), 6-mercaptopurine (CAS no. 50-44- 2), N-(trichloromethylthio)-4-cyclohexane-l,2-dicarboximide (captan) (CAS no. 133- 06-2), 2-aminopurine (CAS no. 452-06-2), methyl methane sulfonate (MMS) (CAS No. 66-27-3), 4-nitroquinoline 1 -oxide (4-NQO) (CAS No. 56-57-5), N4-Aminocytidine (CAS no. 57294-74-3), sodium azide (CAS no. 26628-22-8), N-ethyl-N-nitrosourea (ENU) (CAS no. 759-73-9), N-methyl-N-nitrosourea (MNU) (CAS no. 820-60-0), 5- azacytidine (CAS no. 320-67-2), cumene hydroperoxide (CHP) (CAS no. 80-15-9), ethyl methanesulfonate (EMS) (CAS no. 62-50-0), N-ethyl-N -nitro-N-nitrosoguanidine (ENNG) (CAS no. 4245-77-6), N-methyl-N -nitro-N-nitrosoguanidine (MNNG) (CAS no. 70-25-7), 5-diazouracil (CAS no. 2435-76-9), or t-butyl hydroperoxide (BHP) (CAS no. 75-91-2).

59. The method of any of claims 1-58, wherein the method comprises a phase of diversifying the population of vector by mutagenesis, in which the cells are incubated under conditions suitable for mutagenesis of the gene encoding the protease in the absence of stringent selection for vectors having acquired a gain-of-function mutation in the gene encoding the protease.

60. The method of any of claims 1-44, wherein the method comprises a phase of stringent selection for a mutated replication product of the viral vector encoding an evolved protease.

61. A method of generating protease variants that are resistant to a protease inhibitor, the method comprising

(a) contacting a population of host cells with a population of vectors comprising a gene encoding a protease and deficient in at least one gene for the generation of infectious phage particles, wherein

(1) the host cells are amenable to transfer of the vector;

(2) the vector allows for expression of the protease in the host cell, can be replicated by the host cell, and the replicated vector can transfer into a second host cell;

(3) the host cell expresses a gene product encoded by the at least one gene for the generation of infectious phage particles of (a) in response to the activity of the protease, and the level of gene product expression depends on the activity of the protease;

(b) incubating the population of host cells in the presence of a protease inhibitor and under conditions allowing for mutation of the gene encoding the protease and the transfer of the vector comprising the gene encoding the protease of interest from host cell to host cell, wherein host cells are removed from the host cell population, and the population of host cells is replenished with fresh host cells that do not harbor the vector;

(c) isolating a replicated vector from the host cell population in (b), wherein the replicated vector encodes a mutated protease variant that exhibits an increased protease activity in the presence of the protease inhibitor as compared to the version of the protease of (a).

62. The method of claim 61, further comprising testing a plurality of individual protease inhibitors in separate experiments under identical conditions, and identifying the protease inhibitor to which no resistant protease variant could be identified.

63. The method of claim 61, further comprising testing a plurality of individual protease inhibitors in separate experiments under identical conditions, and identifying the protease inhibitor that required the highest number of mutations in the protease to create a resistant protease variant.

64. The method of any one of claims 61-63, wherein the host cell expresses a fusion molecule comprising

(i) a transcriptional activator; and

(ii) an inhibitor of the transcriptional activator of (i), wherein the inhibitor is fused to the transcriptional activator of (i) via a linker comprising a protease cleavage site that is cleaved by the protease of (a).

65. The method of claim 64, wherein the at least one gene for the generation of infectious phage particles is expressed under the control of a promoter activated by the transcriptional activator of (i).

66. The method of any one of claims 64-65, wherein the inhibitor of (ii) is fused to the N-terminus of the transcriptional activator of (i).

67. The method of any one of claims 64-66, wherein the inhibitor of (ii) is fused to the C-terminus of the transcriptional activator of (i).

68. The method of any one of claims 64-67, wherein the transcriptional activity of the fusion protein is inhibited as compared to the activity of the transcriptional activator alone.

69. The method of any one of claims 61-68, wherein the protease is a therapeutic target.

70. The method of any one of claims 61-69, wherein the protease inhibitor is a therapeutic agent.

71. The method of any one of claims 61-70, wherein the protease inhibitor is used at a concentration that does not completely inhibit the activity of the protease.

72. The method of any one of claims 61-71, wherein the protease inhibitor is used at a concentration that results in an inhibition of at least 50%, at least 75%, at least 80%, at least 90%, at least 95%, at least 97.5%, at least 98%, at least 99%, at least 99.5%, or at least 99.9% of the protease activity.

73. The method of any one of claims 61-72, wherein the protease inhibitor is used at an initial concentration that does not completely inhibit the activity of the protease, and wherein the concentration of the inhibitor is subsequently raised to increase the level of protease inhibition.

74. The method of any one of claims 61-73, wherein the vector is a viral vector.

75. The method of claim 74, wherein the viral vector is a phage.

76. The method of claim 75, wherein the phage is a filamentous phage.

77. The method of claim 76, wherein the phage is an M13 phage.

78. The method of any one of claims 61-77, wherein the protease inhibitor is a protease inhibitor that is in or is approved for clinical use.

79. The method of any one of claims 61-77, wherein the protease inhibitor is a protease inhibitor that is in pre-clinical evaluation.

80. A fusion protein, comprising

(a) a transcriptional activator; and

(b) an inhibitor of the transcriptional activator of (a), wherein the inhibitor is fused to the transcriptional activator of (a) via a linker comprising a protease cleavage site.

81. The fusion protein of claim 80, wherein the inhibitor of (ii) is fused to the N-terminus of the transcriptional activator of (i).

82. The fusion protein of any one of claims 80-81, wherein the inhibitor of (ii) is fused to the C-terminus of the transcriptional activator of (i).

83. The fusion protein of any one of claims 80-83, wherein the transcriptional activity of the fusion protein is inhibited as compared to the activity of the transcriptional activator alone.

84. The fusion protein of any one of claims 80-83, wherein the transcriptional activity of the fusion protein is less than 50%, less than 25%, less than 20%, less than 10%, less than 5%, less than 4%, less than 3%, less than 1%, less than 0.1%, less than 0.01%, or less than 0.001% the activity of the transcriptional activator alone.

85. The fusion protein of any one of claims 83-84, wherein the transcriptional activity of the fusion protein is assessed in a fluorescent or bioluminescent assay.

86. The fusion protein of any one of claims 80-85, wherein cleavage of the protease cleavage site results in an activation of the transcriptional activator.

87. The fusion protein of any one of claims 80-86, wherein the activity of the transcriptional activator comprised in the fusion protein is increased at least2-fold, at least5-fold, at least10-fold, at least20-fold, at least25-fold, at least 50-fold, at least 75-fold, at least 100-fold, at least 200-foled, at least 250-fold, at least 500-fold, at least 750-fold, at least 1000-fold, at least2000-fold, at least2500-fold, at least 5000-fold, at least 7500-fold, or at least 10000-fold upon cleavage of the protease cleavage site of the linker.

88. The fusion protein of any one of claims 80-87, wherein the linker comprises a stretch of at least two consecutive glycine residues on each side of the protease cleavage site.

89. A nucleic acid construct encoding the fusion protein of any one of claims 80-88.

90. A nucleic acid construct, comprising

(a) a nucleic acid sequence encoding a transcriptional activator;

(b) a nucleic acid sequence encoding an inhibitor of the of the transcriptional activator of (a); and

(c) a nucleic acid sequence separating the nucleic acid sequences of (a) and (b), wherein the nucleic acid sequence of (c) encodes a linker and comprises a multiple cloning site allowing for the insertion of a nucleic acid sequence encoding a protease cleavage site.

91. The nucleic acid construct of claim 90, wherein the nucleic acid sequence of (c) further encodes a linker.

92. The nucleic acid construct of claim 91, wherein the linker comprises a stretch of at least two consecutive glycine residues on each side of the protease cleavage site.

93. A mutagenesis plasmid comprising a gene expression cassette encoding a component of *E. coli* translesion synthesis polymerase V, a deoxyadenosine methylase, and/or a hemimethylated-GATC binding domain, or any combination thereof.

94. The mutagenesis plasmid of claim 93, wherein the component of *E. coli* translesion synthesis polymerase V is *umuC*.

95. The mutagenesis plasmid of claim 93 or 94, wherein the deoxyadenosine methylase is *dam*.

96. The mutagenesis plasmid of any one of claims 93-95, wherein the hemimethylated-GATC binding domain is *seqA*.

97. A kit comprising
(a) a vector encoding an M13 phage backbone and a multiple cloning site for insertion of a nucleic acid sequence encoding a protease, wherein the vector or a replication product thereof can be packaged into infectious phage particles in the presence of other phage functions by suitable host cells, but lacks at least one gene required for the generation of infectious particles;
(b) an accessory plasmid comprising a nucleic acid sequence encoding the at least one gene of interest under the control of a promoter that is activated by a transcriptional activator; and
(c) an expression construct encoding a fusion protein of the transcriptional activator that activates the promoter of (b) fused to an inhibitor of the transcriptional activator via a linker, and a multiple cloning site for insertion of a nucleic acid sequence encoding a protease cleavage site.

98. The kit of claim 97, wherein the kit further comprises a helper phage providing all phage functions except for the at least one gene required for the generation of infectious phage particles provided by the accessory plasmid of (b).

99. The kit of claim 97 or 98, wherein the helper phage or a replication product thereof cannot be packaged into infectious phage particles.

100. The kit of any one of claims 97-99, wherein the kit comprises suitable host cells.

101. The kit of any one of claims 97-100, wherein the host cells are E. coli host cells.

102. The kit of any one of claims 97-101, wherein the kit further comprises a mutagenesis plasmid.

103. The kit of claim 102, wherein the mutagenesis plasmid comprising a gene expression cassette encoding *umuC* (a components of E. coli translesion synthesis polymerase V), *dam* (deoxyadenosine methylase), and/or *seqA* (a hemimethylated-GATC binding domain), or any combination thereof.

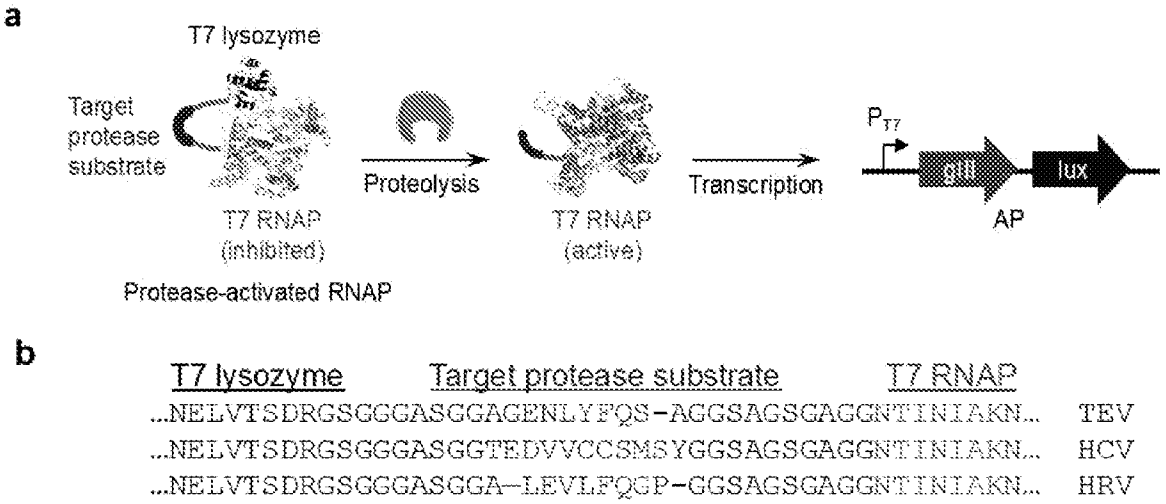
104. A protease generated by the method of any one of claims 1-79, or comprising at least one mutation observed in a protease generated by the method of any one of claims 1-79, wherein the protease is not a wild-type or naturally occurring protease.

105. A method comprising contacting a protein with a protease generated by the method of any one of claims 1-79 or with the protease of claim 104 under conditions suitable for the protease to cleave the protein.

106. The method of claim 105, wherein the contacting is *in vivo*.

107. The method of claim 105 or 106, wherein the protein is associated with a disease or disorder.

108. The method of any one of claims 105-106, wherein the protein is C-C chemokine receptor type 5 (CCR5), Programmed death-ligand 1 (PDL1), Tumor Necrosis Factor alpha (TNFa), Insulin-Degrading Enzyme (IDE), membrane metallo-endopeptidase (MME, Neprilysin), or Interleukin 23 alpha subunit P19 (IL23aP19).



b

T7 lysozyme	Target protease substrate	T7 RNAP	
...NELVTSDRGSGGGASGGAGENLYFQS	-AGGSAGSGAGGNTINIAKN...		TEV
...NELVTSDRGSGGGASGGTEDVVCCSMSY	GGSAGSGAGGNTINIAKN...		HCV
...NELVTSDRGSGGGASGGA—LEVLFQGP	-GGSAGSGAGGNTINIAKN...		HRV

FIG. 1

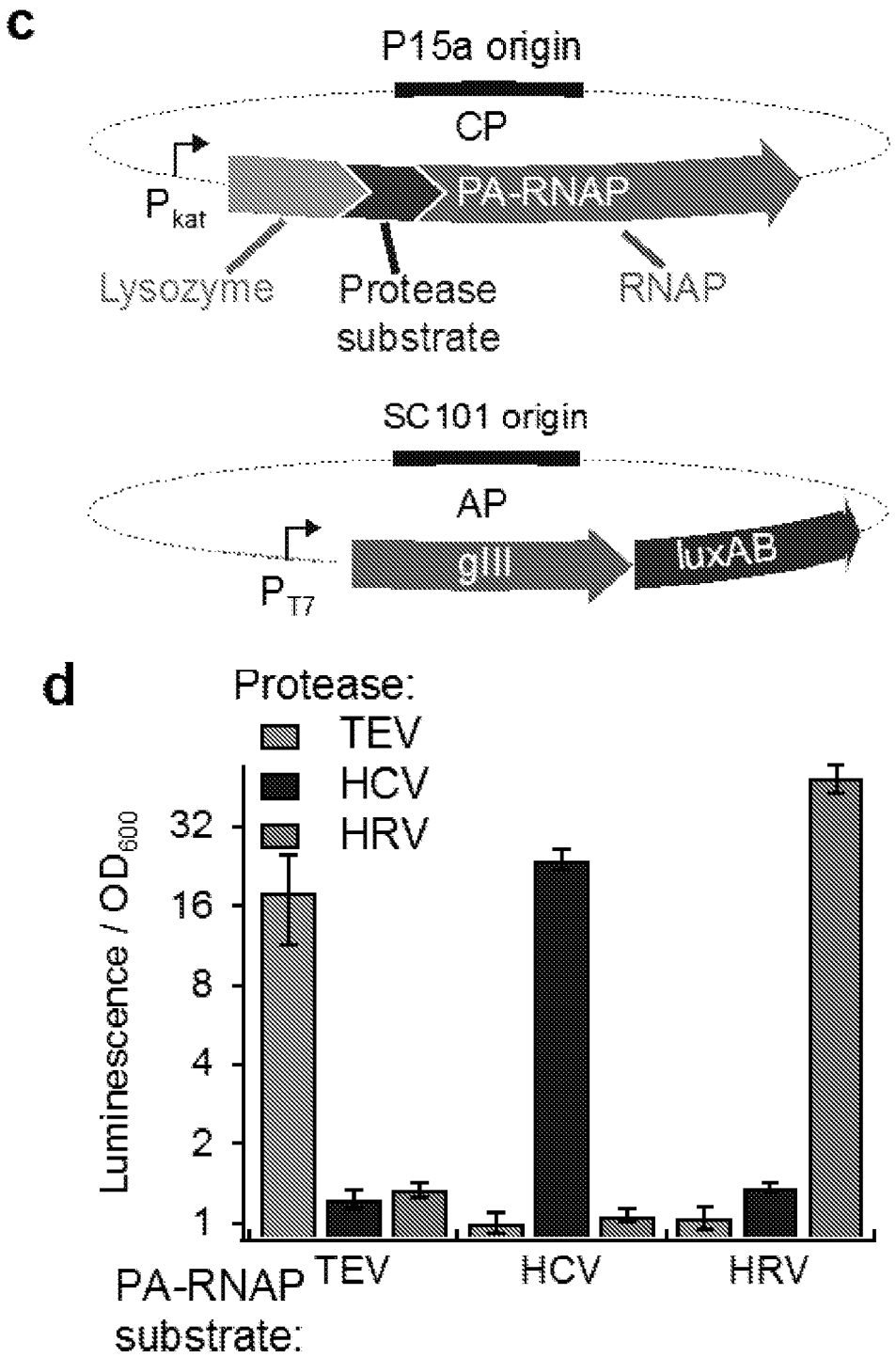


FIG. 1 (continued)

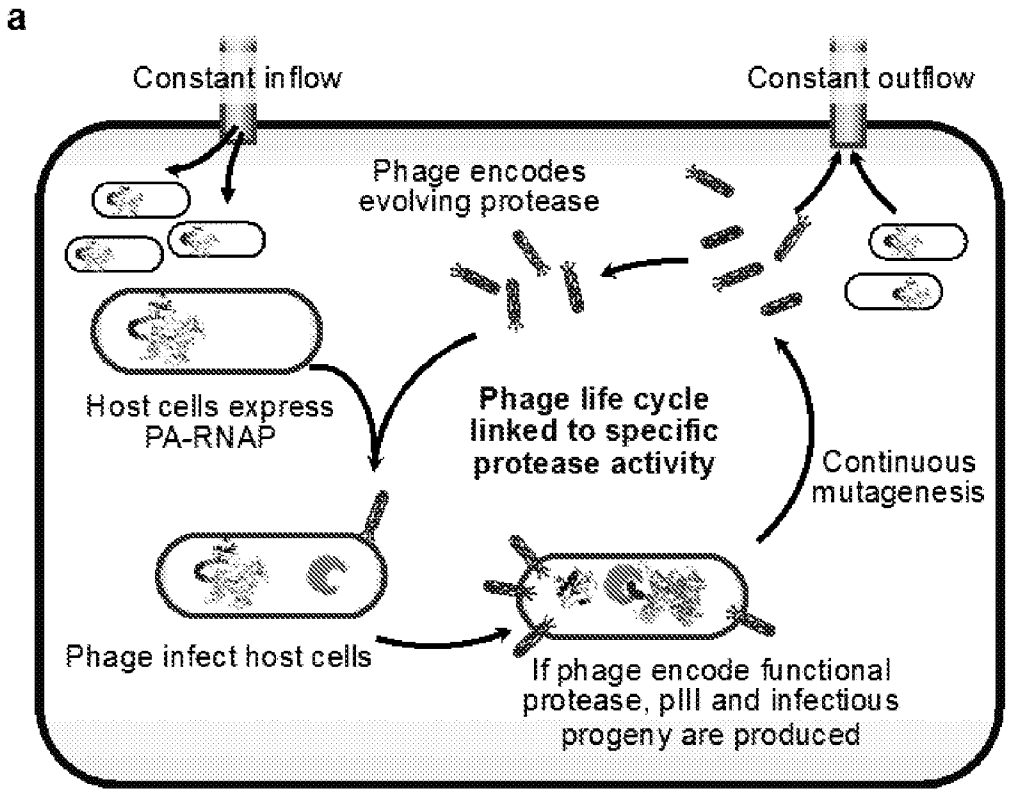

a
Constant inflow
Constant outflow
Phage encodes
evolving protease
Host cells express
PA-RNAP
Phage life cycle
linked to specific
protease activity
Continuous
mutagenesis
Phage infect host cells
If phage encode functional
protease, pIII and infectious
progeny are produced

FIG. 2

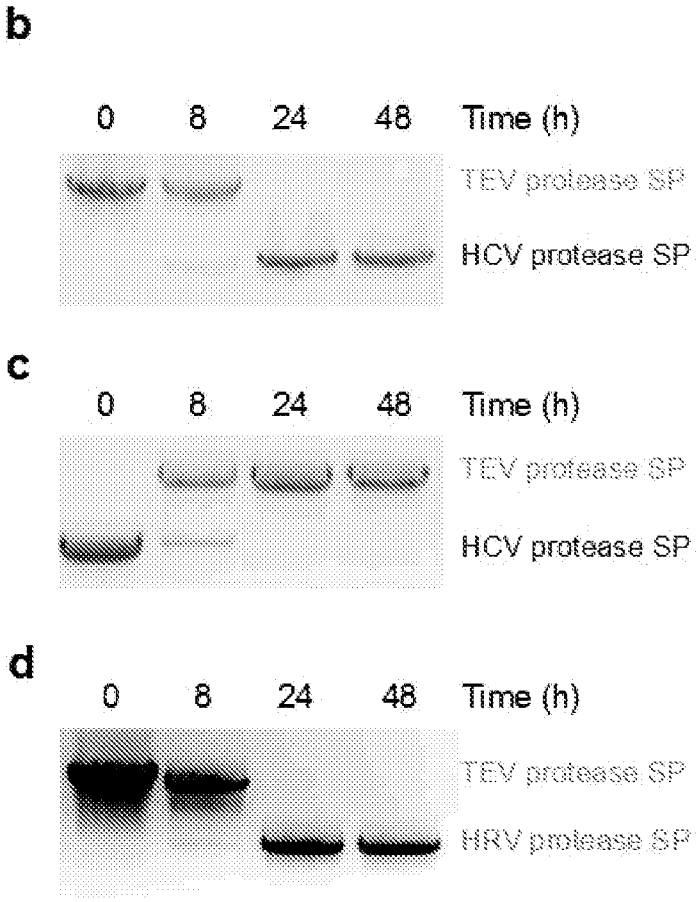


FIG. 2 (continued)

a

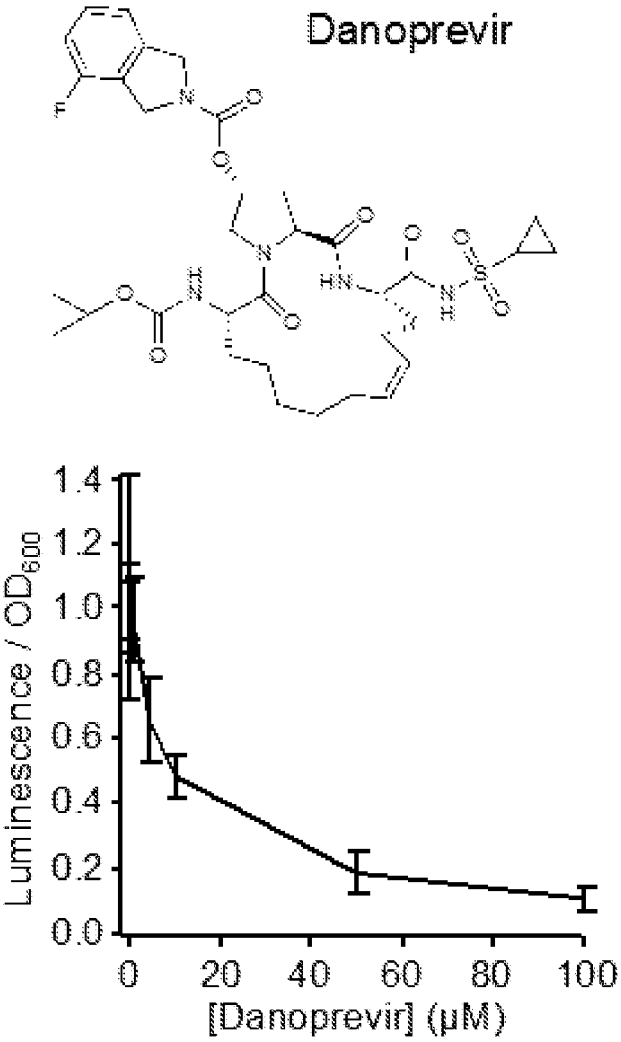

Danoprevir
Luminescence / OD600
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
[Danoprevir] (μM)

FIG. 3

b

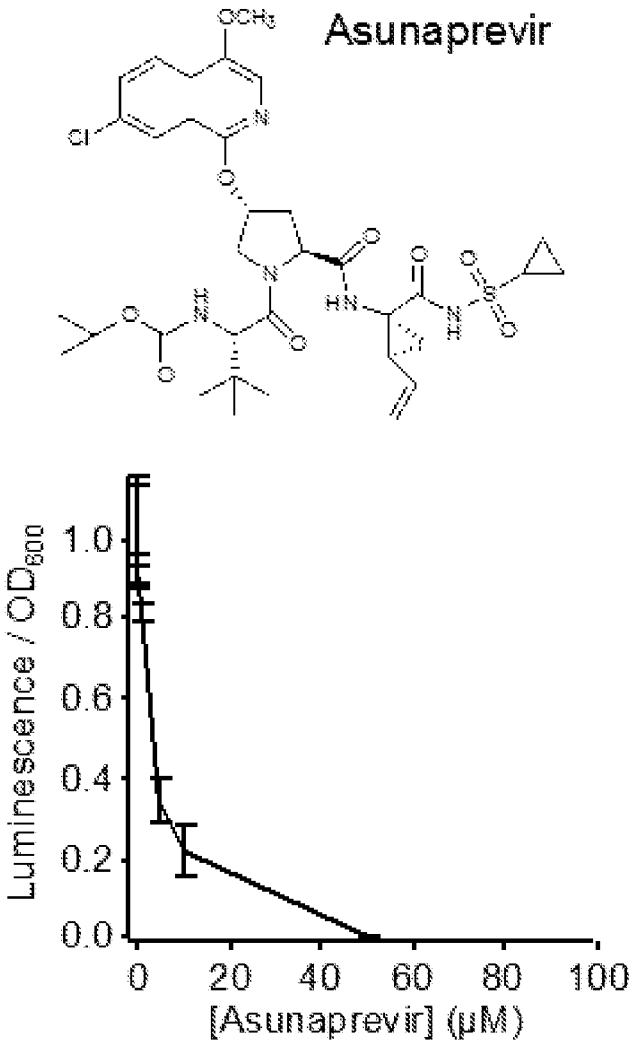


FIG. 3 (continued)

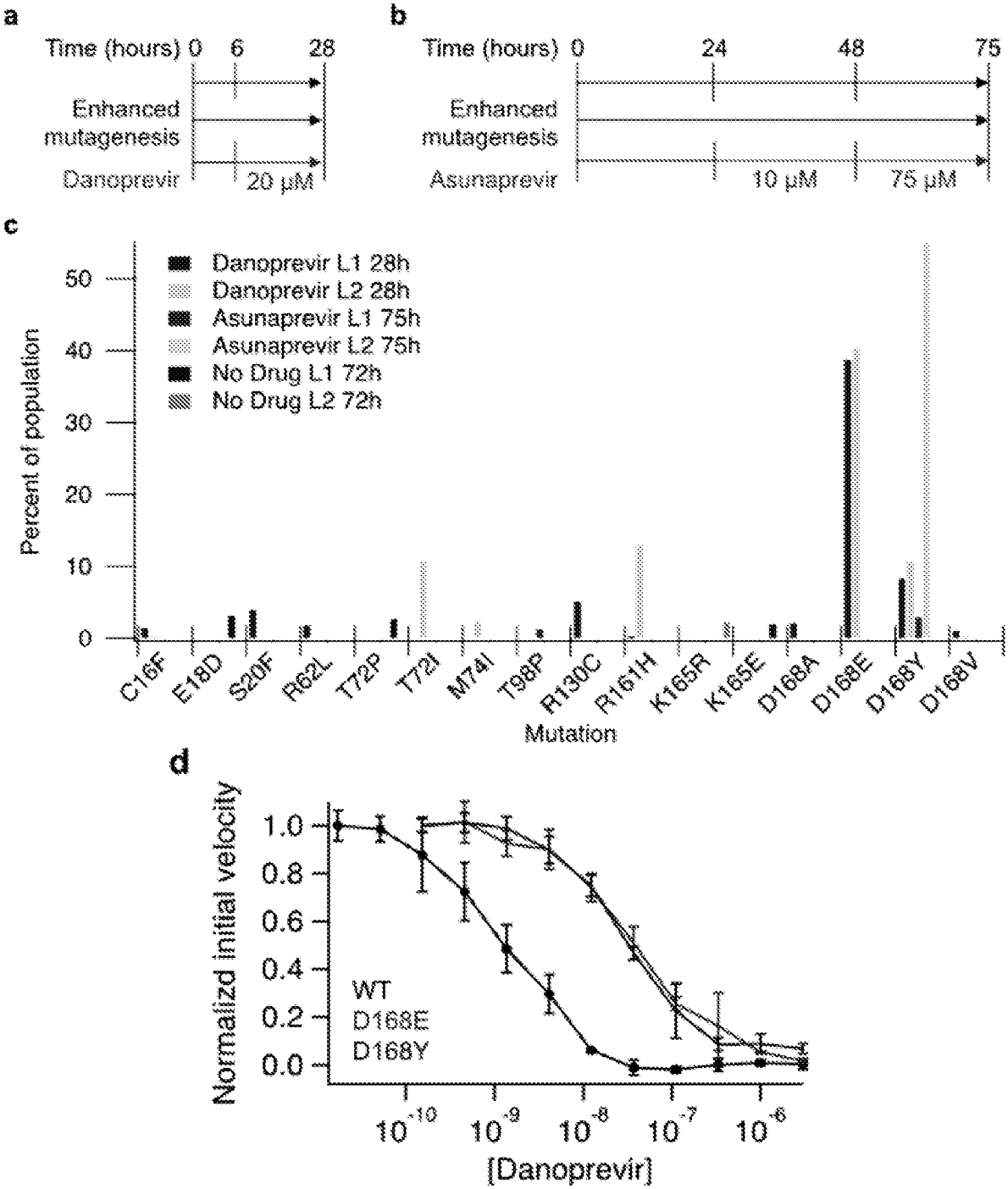

a
Time (hours) 0 6 28
Enhanced mutagenesis
Danoprevir
20 μM
b
Time (hours) 0 24 48 75
Enhanced mutagenesis
Asunaprevir
10 μM
75 μM
c
Danoprevir L1 28h
Danoprevir L2 28h
Asunaprevir L1 75h
Asunaprevir L2 75h
No Drug L1 72h
No Drug L2 72h
Percent of population
0
10
20
30
40
50
C16F
E18D
S20F
R62L
T72P
T72I
M74I
T98P
R130C
R161H
K165R
K165E
D168A
D168E
D168Y
D168V
Mutation
d
Normalized initial velocity
1.0
0.8
0.6
0.4
0.2
0.0
WT
D168E
D168Y
10^-10
10^-9
10^-8
10^-7
10^-6
[Danoprevir]

FIG. 4

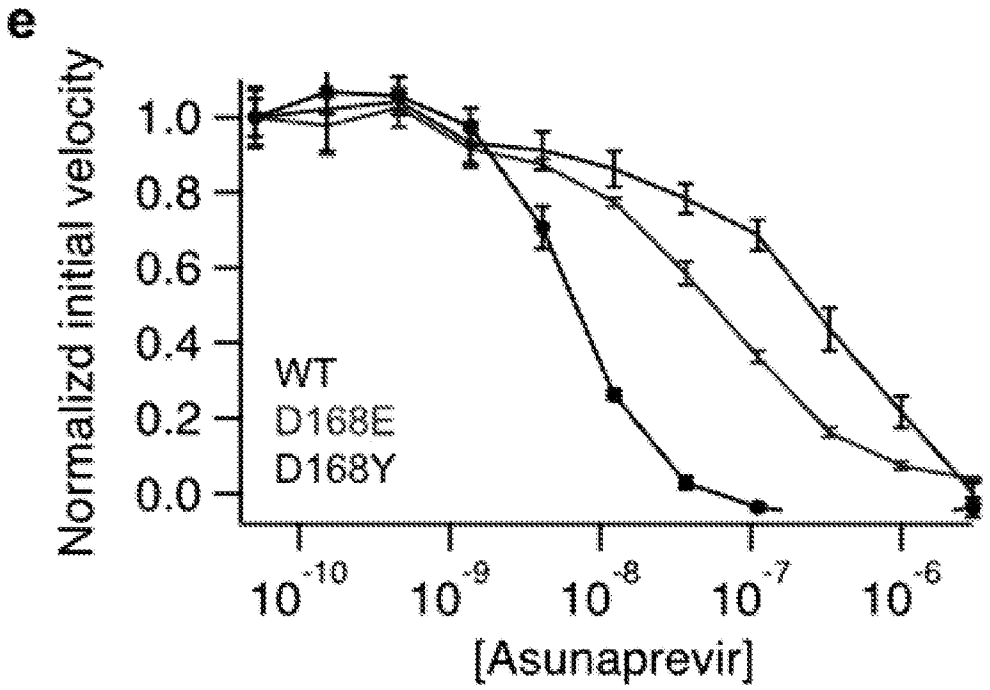


FIG. 4 Continued

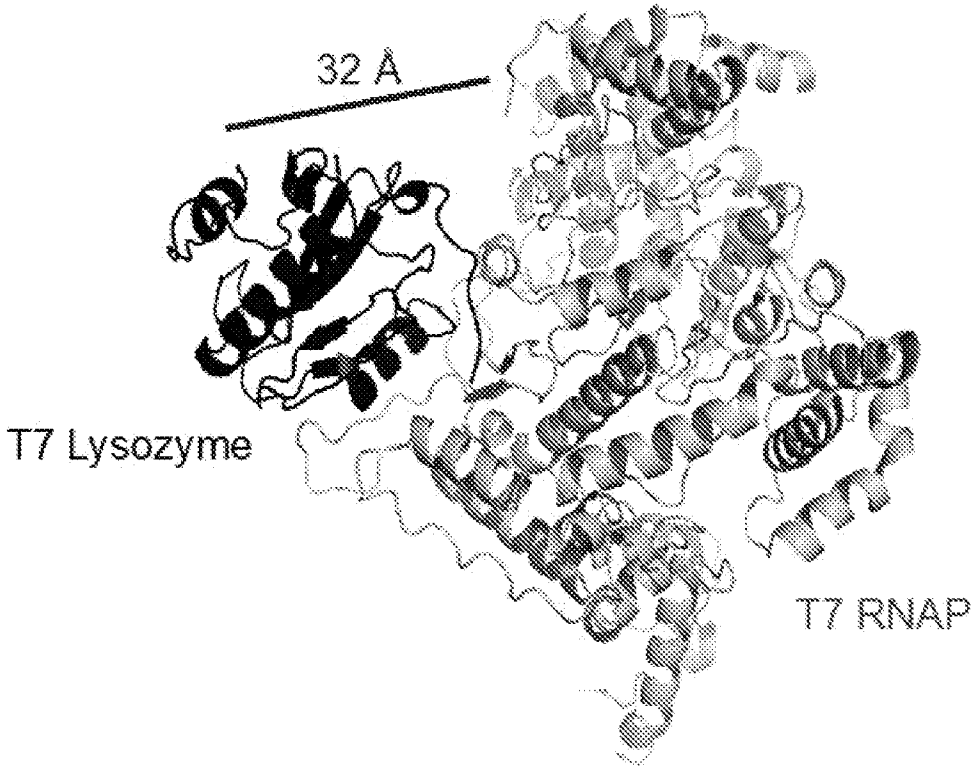


FIG. 5

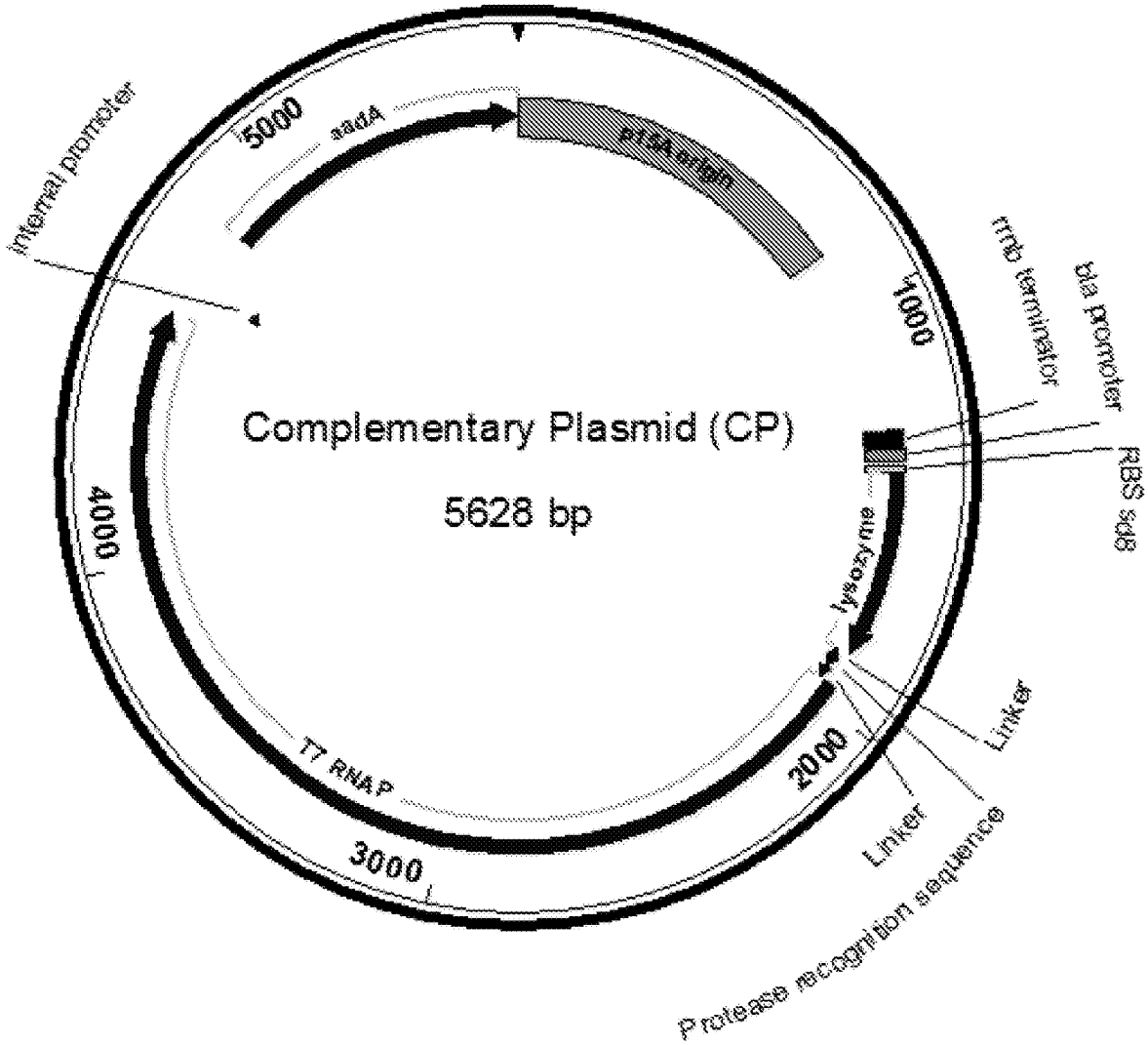

Complementary Plasmid (CP)
5628 bp
5000
aadA
P15A origin
1000
rrnb terminator
bla promoter
RBS sgB
lysozyme
2000
Linker
Linker
Protease recognition sequence
3000
T7 RNAP
4000
Internal promoter

FIG. 6

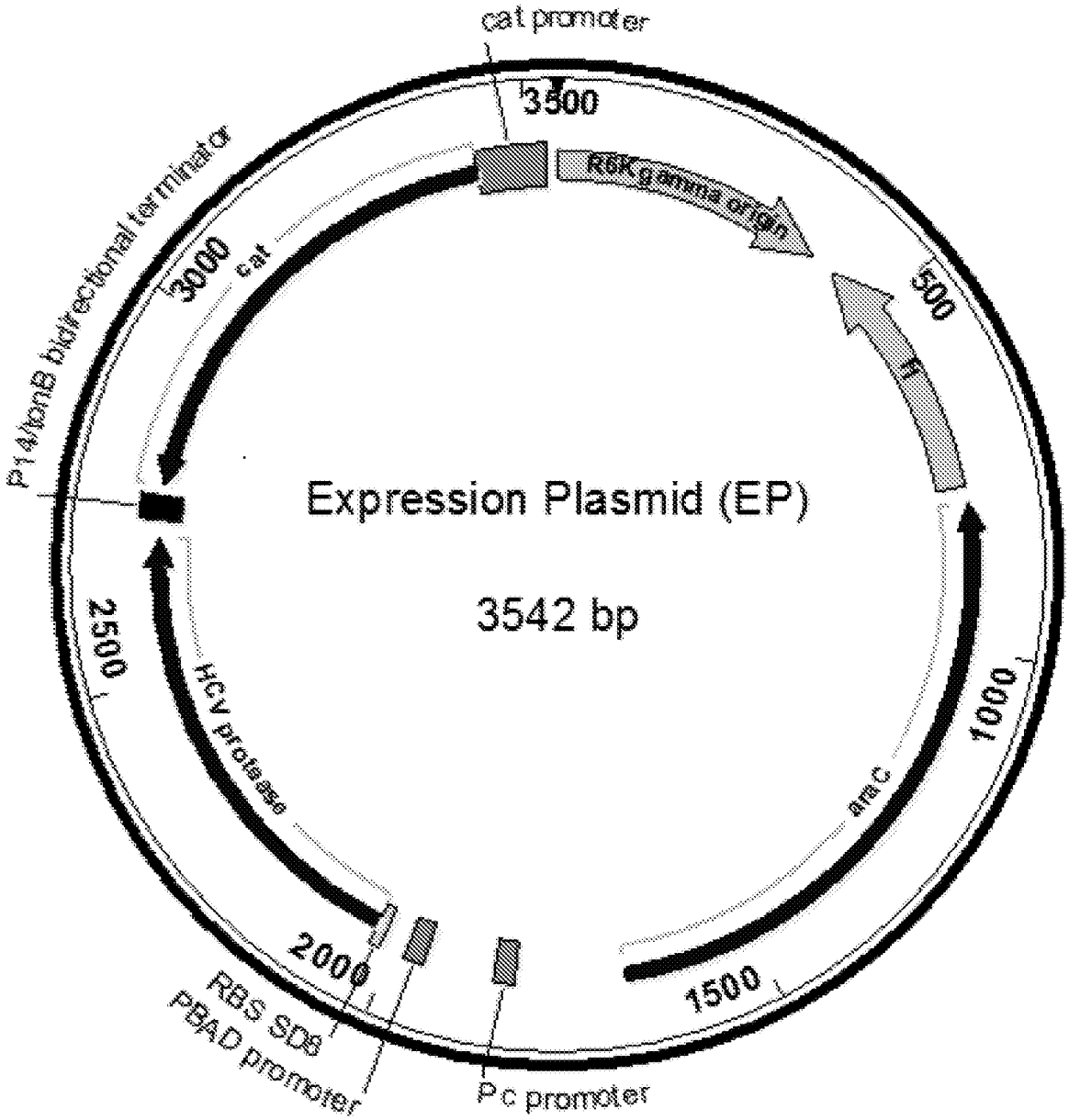

cat promoter
3500
R6K gamma origin
500
ff
1000
araC
1500
Pc promoter
2000
RBS SD9
PBAD promoter
2500
HCV protease
3000
cat
P14/tonB bidirectional terminator
Expression Plasmid (EP)
3542 bp

FIG. 7

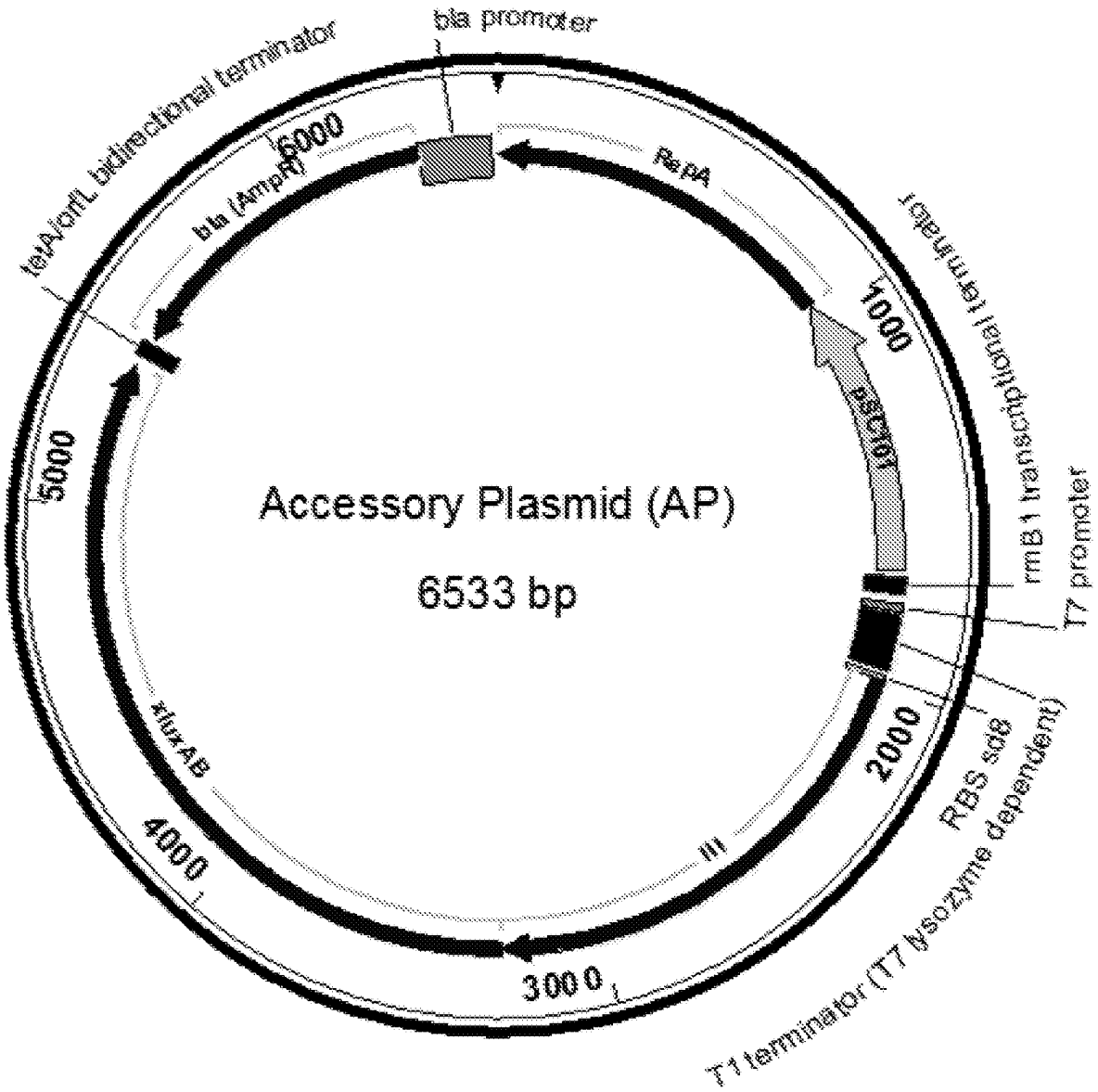

bla promoter
tetA/orfL bidirectional terminator
6000
bla (AmpR)
RepA
1000
pSC101
rmB1 transcriptional terminator
T7 promoter
5000
Accessory Plasmid (AP)
6533 bp
xluxAB
2000
RBS sd8
T1 terminator (T7 lysozyme dependent)
4000
III
3000

FIG. 8

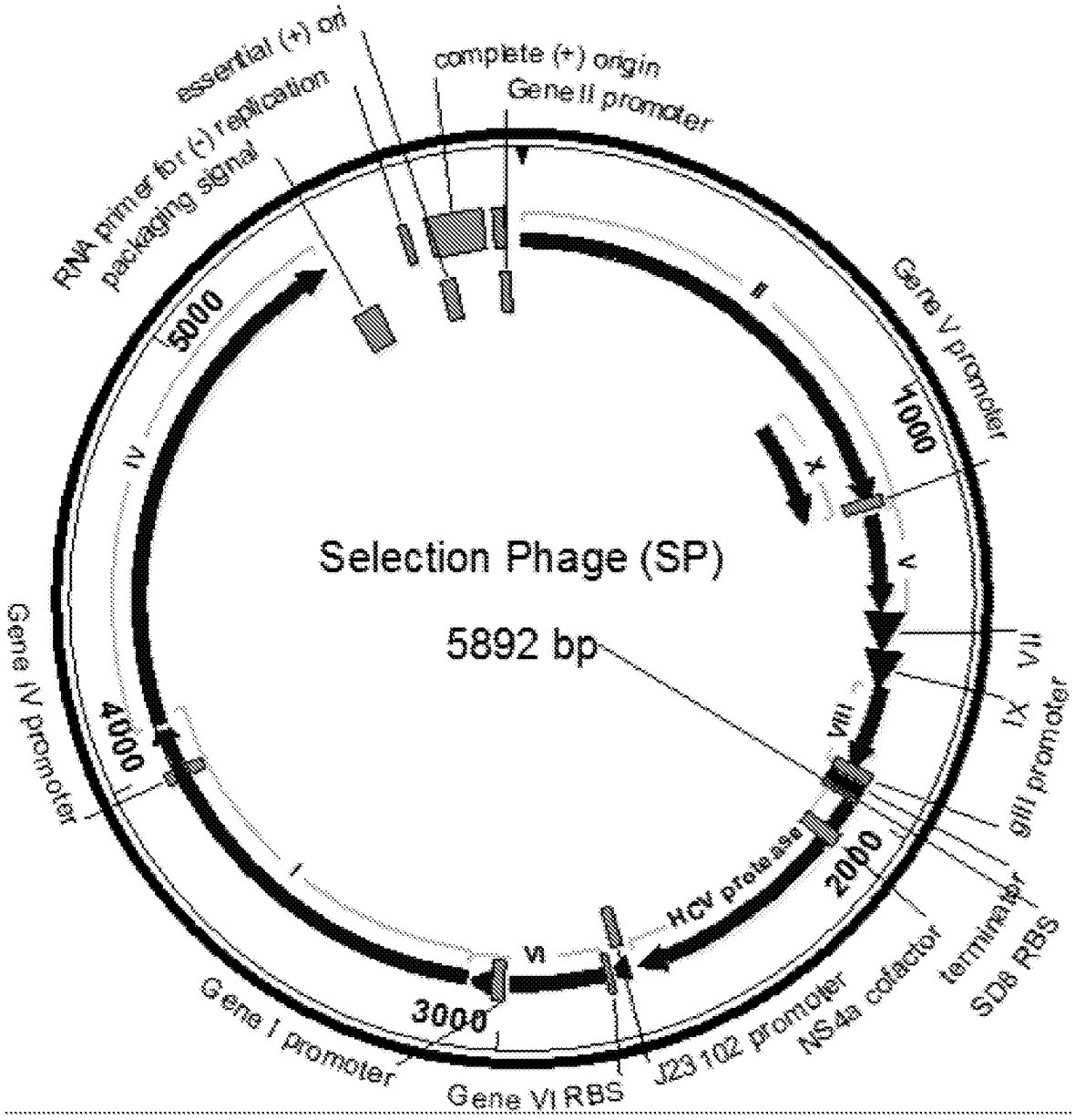

essential (+) ori
complete (+) origin
Gene II promoter
RNA primer for (-) replication
packaging signal
Gene V promoter
1000
5000
II
IV
X
V
Selection Phage (SP)
5892 bp
VII
IX
VIII
gIII promoter
Gene IV promoter
4000
2000
HCV protease
I
VI
terminator
SD8 RBS
Gene I promoter
3000
NS4a cofactor
J23102 promoter
Gene VI RBS

FIG. 9

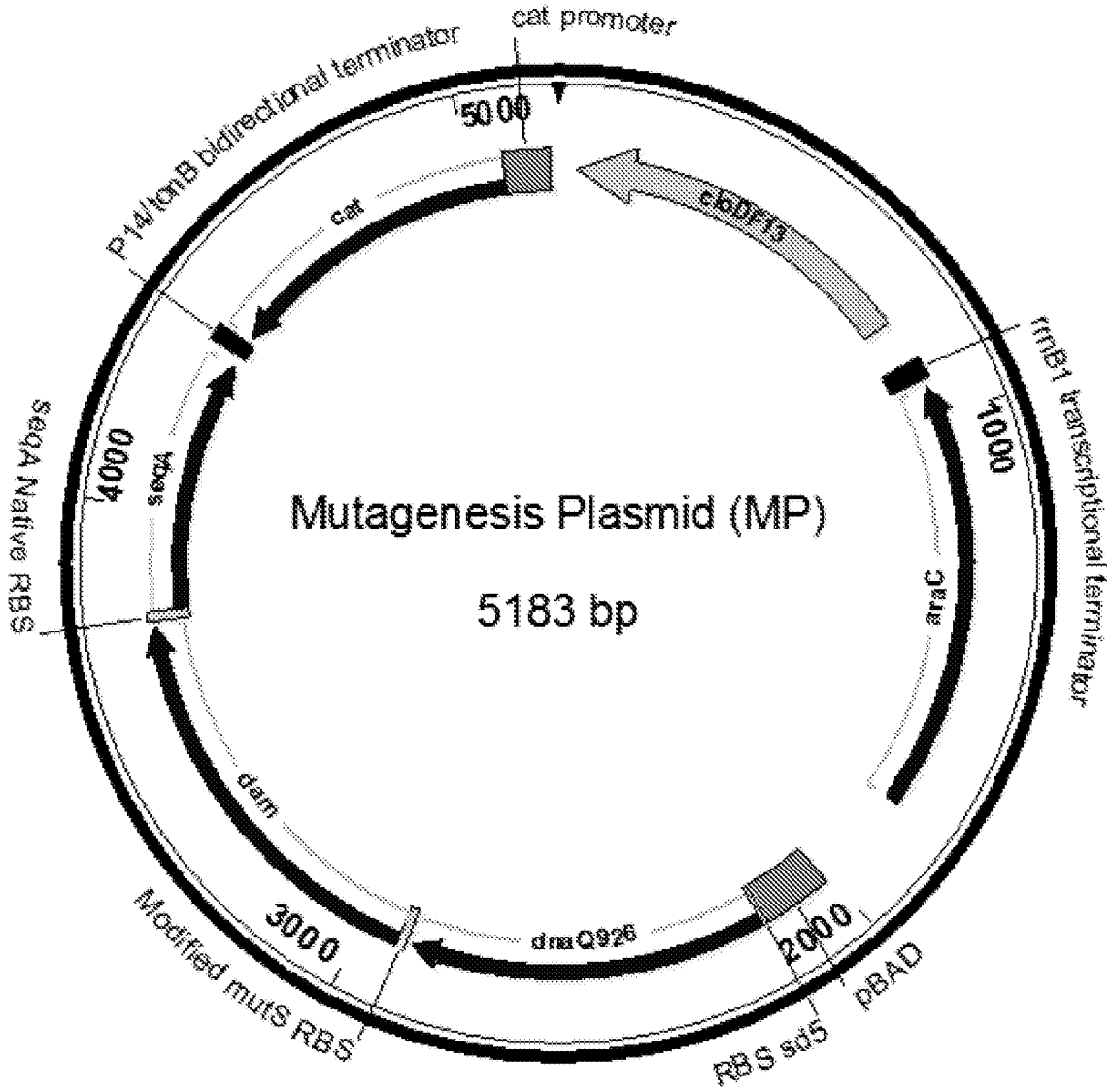

cat promoter
P14/tonB bidirectional terminator
5000
cat
c60F13
rrnB1 transcriptional terminator
1000
seqA Native RBS
4000
seqA
Mutagenesis Plasmid (MP)
5183 bp
araC
dam
3000
dnaQ926
2000
Modified mutS RBS
pBAD
RBS sd5

FIG. 10

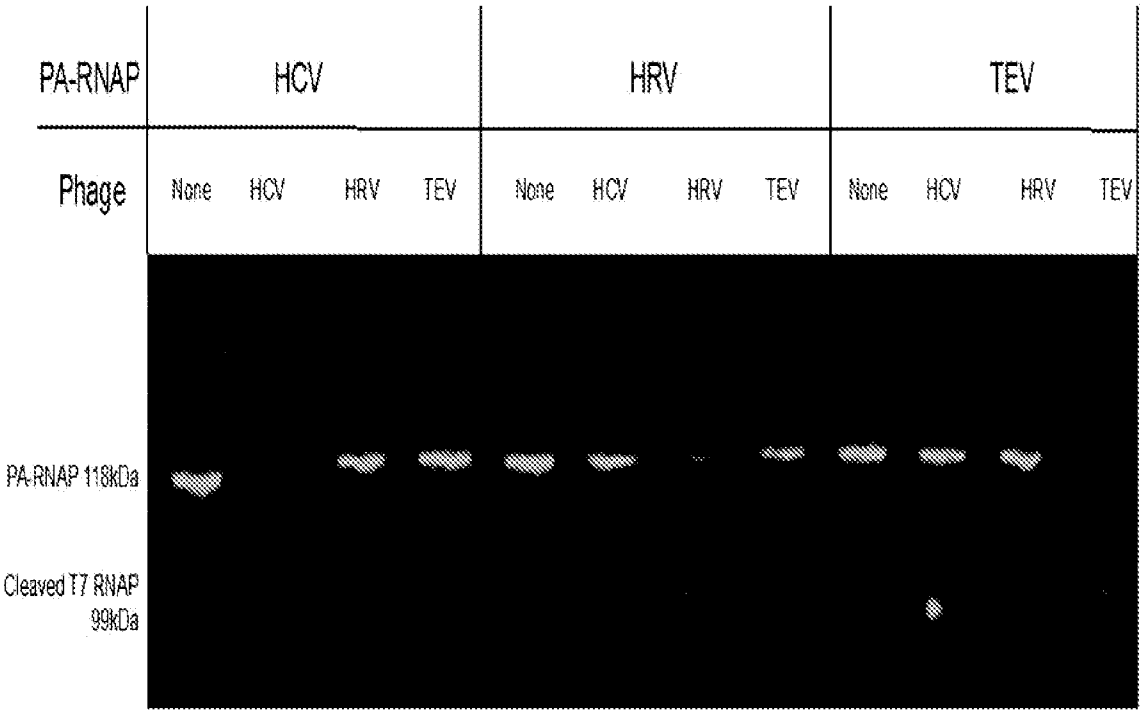

PA-RNAP
HCV
HRV
TEV
Phage
None
HCV
HRV
TEV
None
HCV
HRV
TEV
None
HCV
HRV
TEV
PA-RNAP 118kDa
Cleaved T7 RNAP
99kDa

FIG. 11

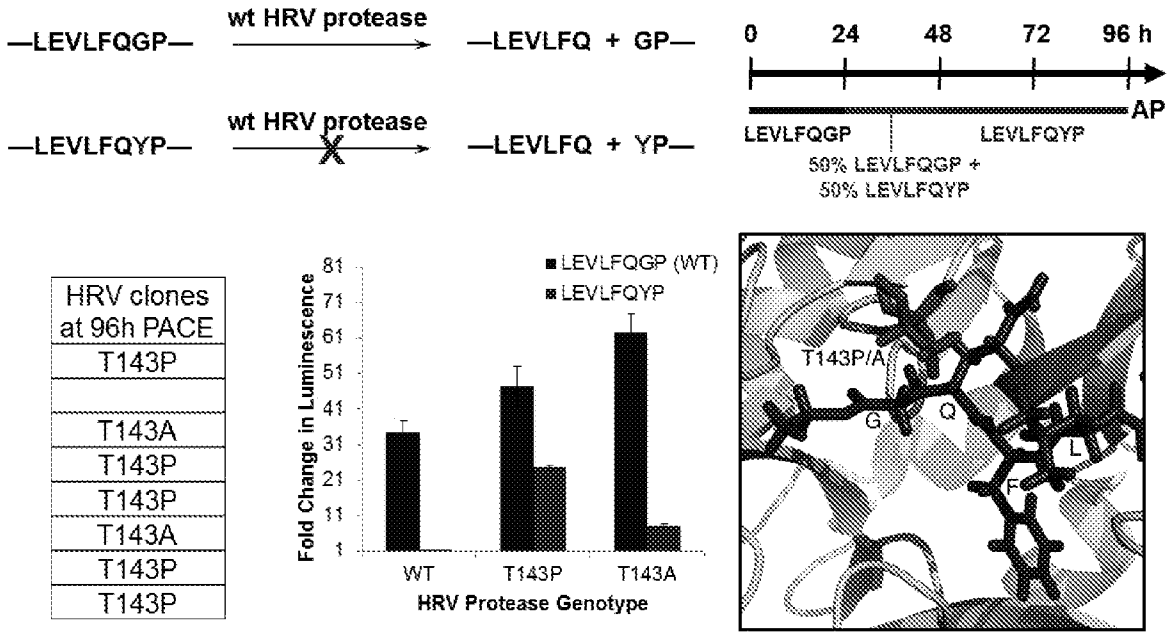

—LEVLFQGP—
wt HRV protease
—LEVLFQ + GP—
—LEVLFQYP—
wt HRV protease
—LEVLFQ + YP—
0
24
48
72
96 h
AP
LEVLFQGP
LEVLFQYP
50% LEVLFQGP +
50% LEVLFQYP
HRV clones at 96h PACE
T143P
T143A
T143P
T143P
T143A
T143P
T143P
LEVLFQGP (WT)
LEVLFQYP
Fold Change in Luminescence
81
71
61
51
41
31
21
11
1
WT
T143P
T143A
HRV Protease Genotype
T143P/A
G
Q
L
F

FIG. 12

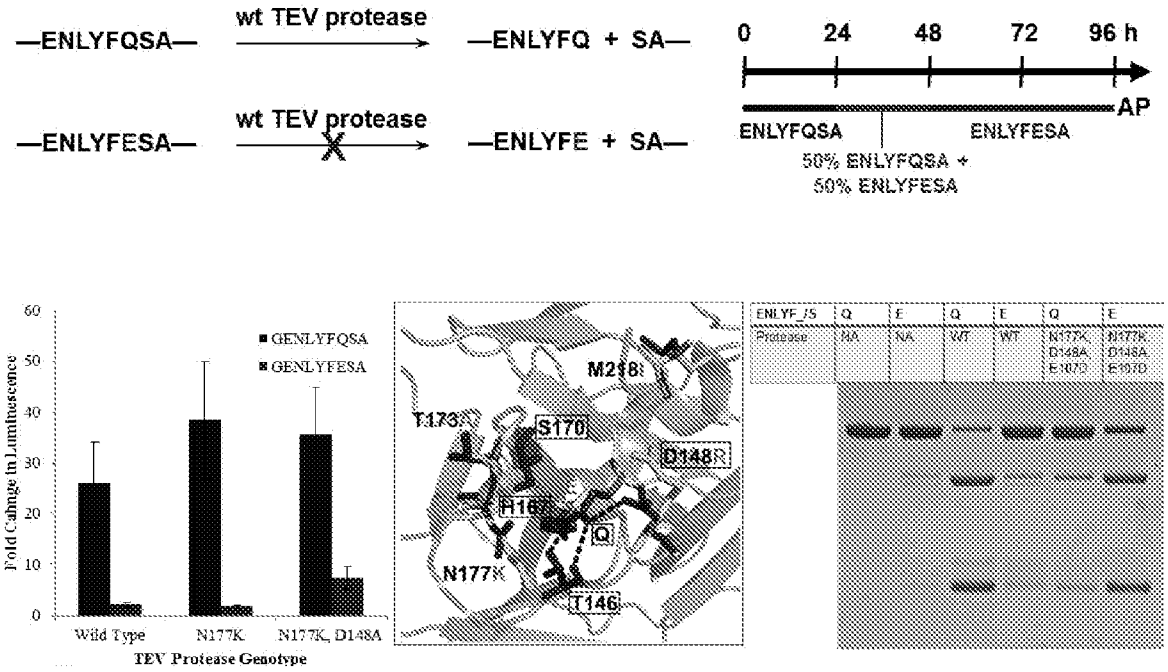
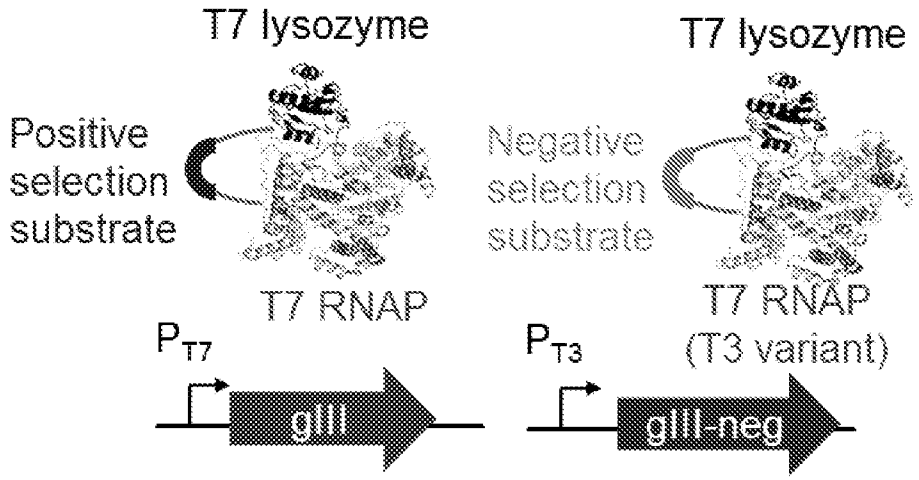
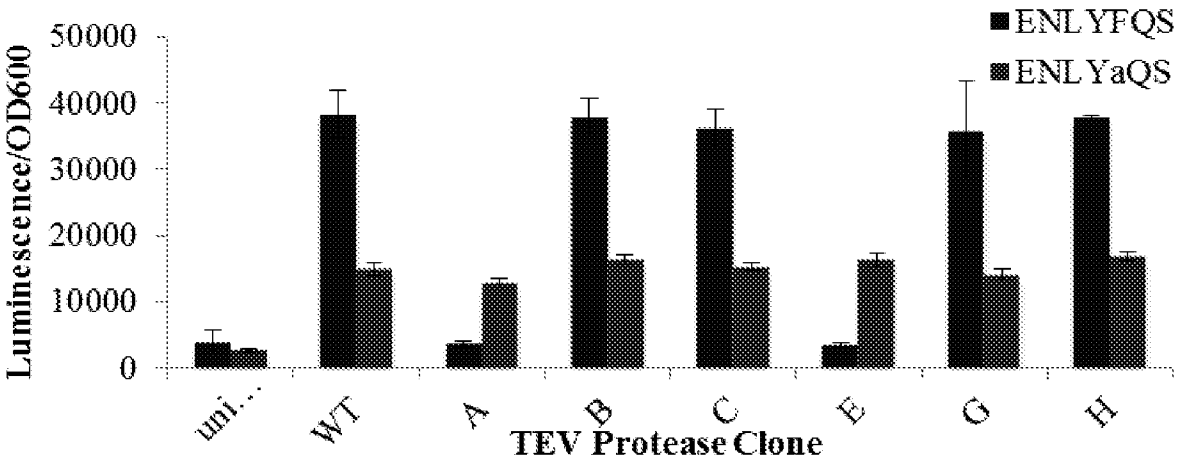

—ENLYFQSA—
wt TEV protease
—ENLYFQ + SA—
—ENLYFESA—
wt TEV protease
—ENLYFE + SA—
0
24
48
72
96 h
AP
ENLYFQSA
ENLYFESA
50% ENLYFQSA +
50% ENLYFESA
GENLYFQSA
GENLYFESA
Fold Change in Luminescence
Wild Type
N177K
N177K, D148A
TEV Protease Genotype
M218
T173
S170
D148
H167
Q
N177
T146
ENLYF/S
Q
E
Q
E
Q
E
Protease
NA
NA
WT
WT
N177K, D148A, E107D
N177K, D148A, E107D

FIG. 13


T7 lysozyme
Positive selection substrate
T7 RNAP
P_{T7}
gIII
T7 lysozyme
Negative selection substrate
T7 RNAP (T3 variant)
P_{T3}
gIII-neg


Luminescence/OD600
50000
40000
30000
20000
10000
0
ENLYFQS
ENLYaQS
uni...
WT
A
B
C
E
G
H
TEV Protease Clone

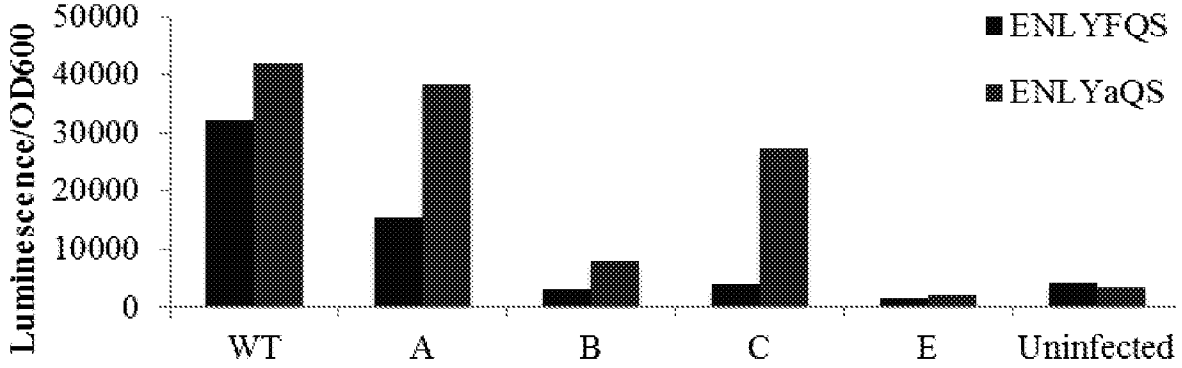

Luminescence/OD600
50000
40000
30000
20000
10000
0
ENLYFQS
ENLYaQS
WT
A
B
C
E
Uninfected

FIG. 14

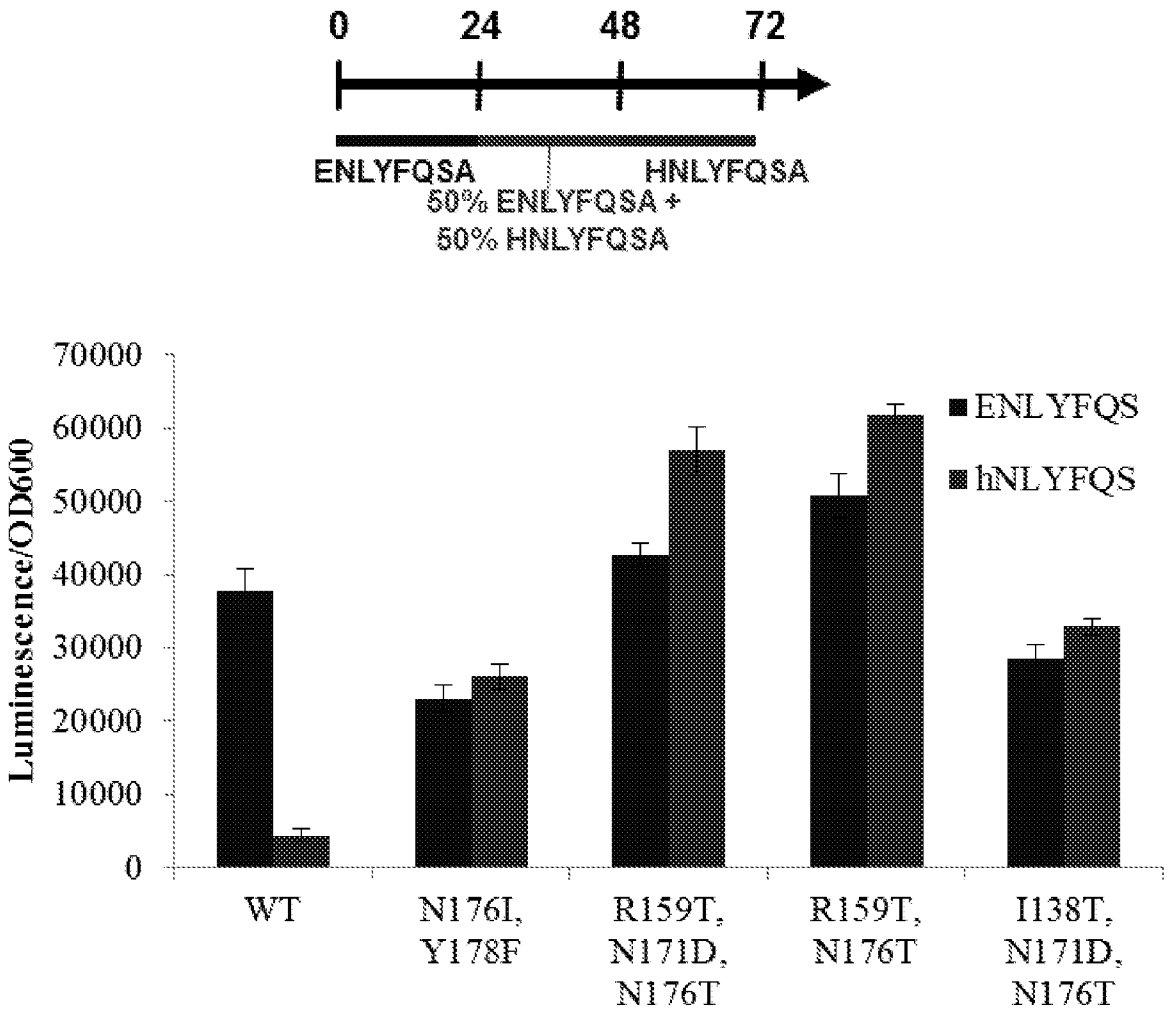

0
24
48
72
ENLYFQSA
HNLYFQSA
50% ENLYFQSA +
50% HNLYFQSA
Luminescence/OD600
70000
60000
50000
40000
30000
20000
10000
0
ENLYFQS
hNLYFQS
WT
N176I, Y178F
R159T, N171D, N176T
R159T, N176T
I138T, N171D, N176T

FIG. 15

Substrate _NLYFQ/S	E	H	E	H	E	H
Protease	NA	NA	WT	WT	I138T, N171D, N176T	I138T, N171D, N176T

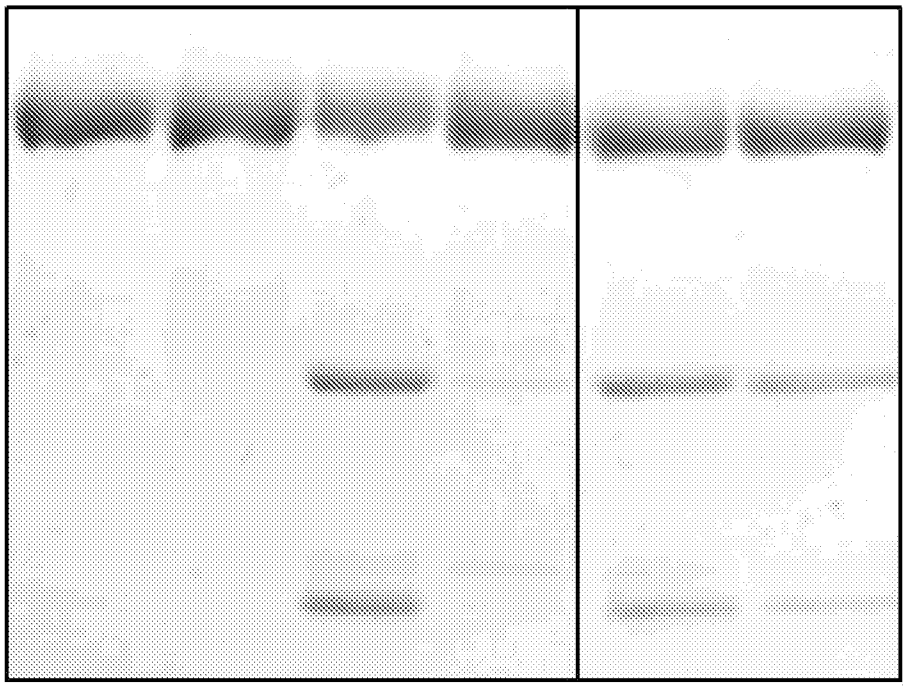


FIG. 16

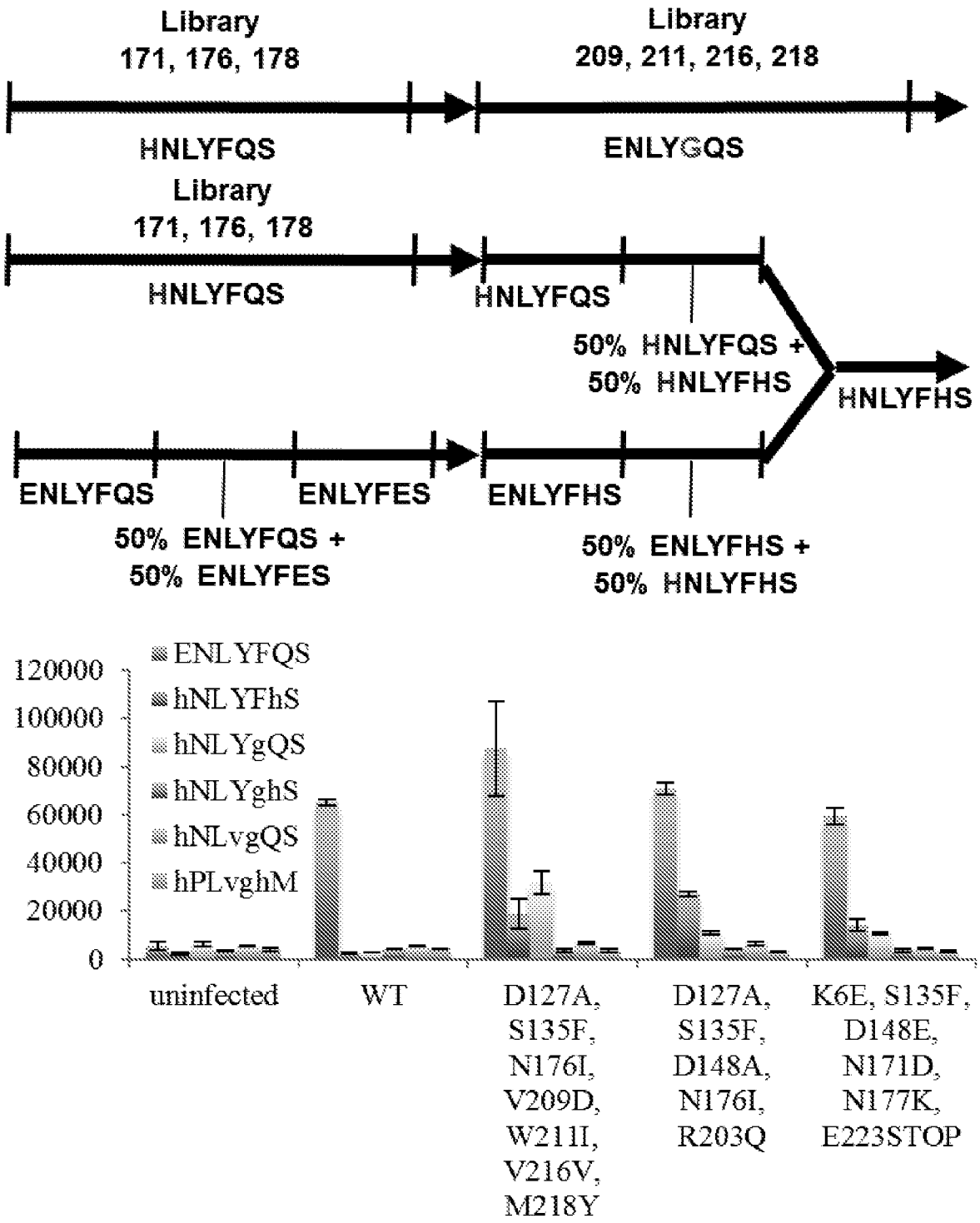

Library
171, 176, 178
HNLYFQS
Library
209, 211, 216, 218
ENLYGQS
Library
171, 176, 178
HNLYFQS
HNLYFQS
50% HNLYFQS +
50% HNLYFHS
HNLYFHS
ENLYFQS
ENLYFES
ENLYFHS
50% ENLYFQS +
50% ENLYFES
50% ENLYFHS +
50% HNLYFHS
ENLYFQS
hNLYFhS
hNLYgQS
hNLYghS
hNLvgQS
hPLvghM
120000
100000
80000
60000
40000
20000
0
uninfected
WT
D127A, S135F, N176I, V209D, W211I, V216V, M218Y
D127A, S135F, D148A, N176I, R203Q
K6E, S135F, D148E, N171D, N177K, E223STOP

FIG. 17

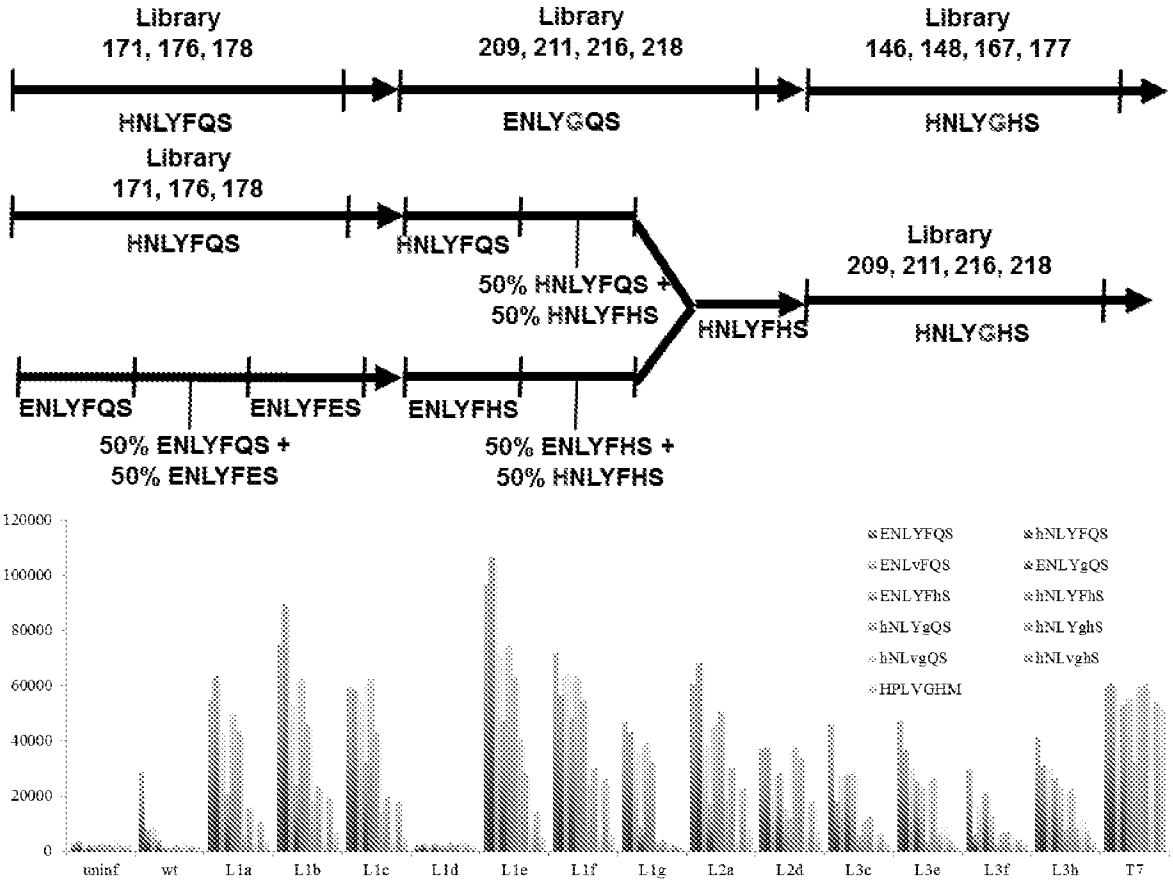
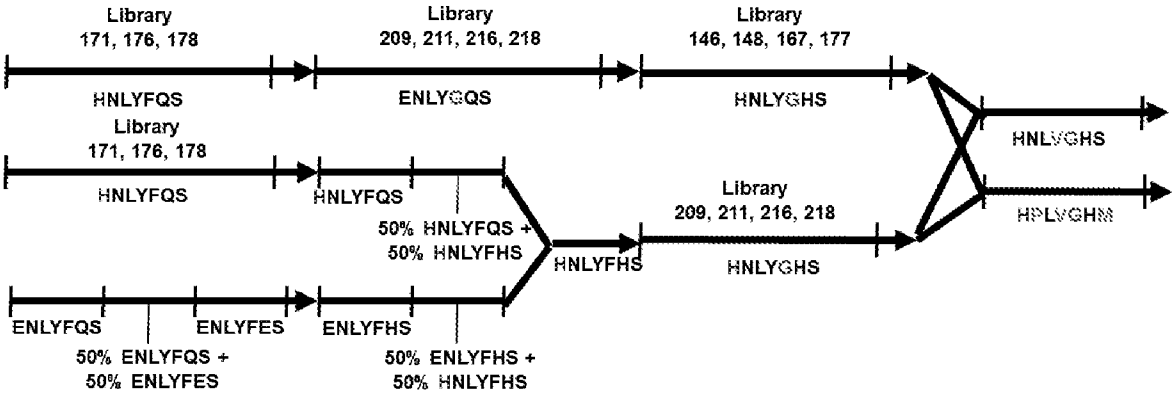

Library
171, 176, 178
HNLYFQS
Library
209, 211, 216, 218
ENLYGQS
Library
146, 148, 167, 177
HNLYGHS
Library
171, 176, 178
HNLYFQS
HNLYFQS
50% HNLYFQS +
50% HNLYFHS
HNLYFHS
Library
209, 211, 216, 218
HNLYGHS
ENLYFQS
ENLYFES
ENLYFHS
50% ENLYFQS +
50% ENLYFES
50% ENLYFHS +
50% HNLYFHS
120000
100000
80000
60000
40000
20000
0
ENLYFQS
ENLvFQS
ENLYFhS
hNLYgQS
hNLvgQS
HPLVGHM
hNLYFQS
ENLYgQS
hNLYFhS
hNLYghS
hNLvghS
uninf
wt
L1a
L1b
L1c
L1d
L1e
L1f
L1g
L2a
L2d
L3c
L3e
L3f
L3h
T7

FIG. 18


Library
171, 176, 178
HNLYFQS
Library
209, 211, 216, 218
ENLYGQS
Library
146, 148, 167, 177
HNLYGHS
HNLVGHS
Library
171, 176, 178
HNLYFQS
HNLYFQS
50% HNLYFQS +
50% HNLYFHS
HNLYFHS
Library
209, 211, 216, 218
HNLYGHS
HPLVGHM
ENLYFQS
ENLYFES
ENLYFHS
50% ENLYFQS +
50% ENLYFES
50% ENLYFHS +
50% HNLYFHS

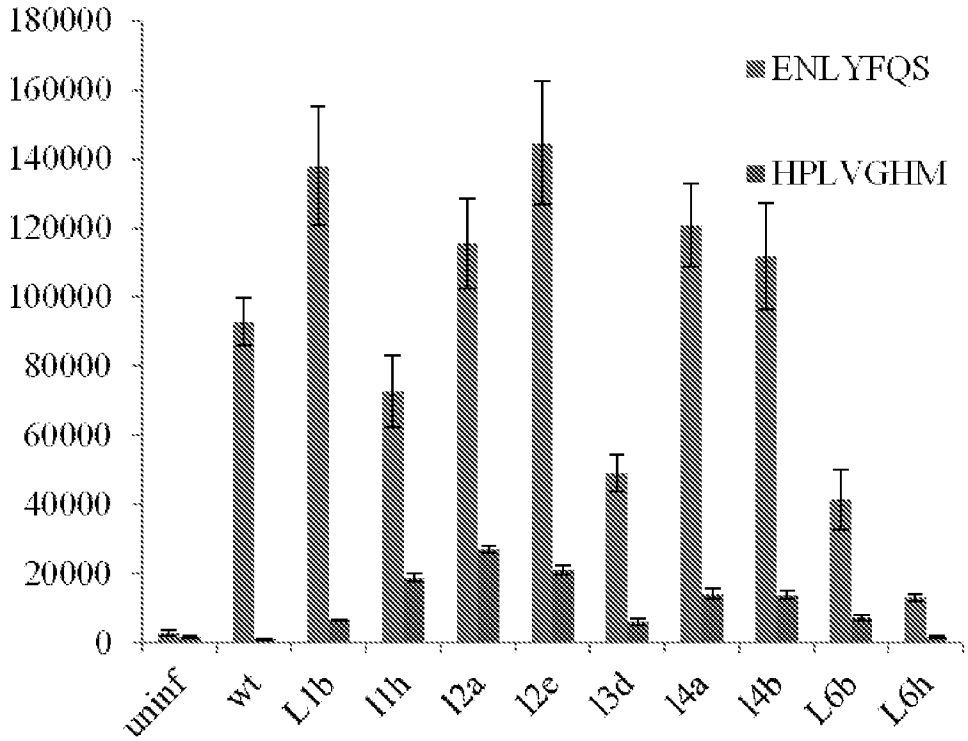

180000
160000
140000
120000
100000
80000
60000
40000
20000
0
ENLYFQS
HPLVGHM
uninf
wt
L1b
l1h
l2a
l2e
l3d
l4a
l4b
L6b
L6h

FIG. 19

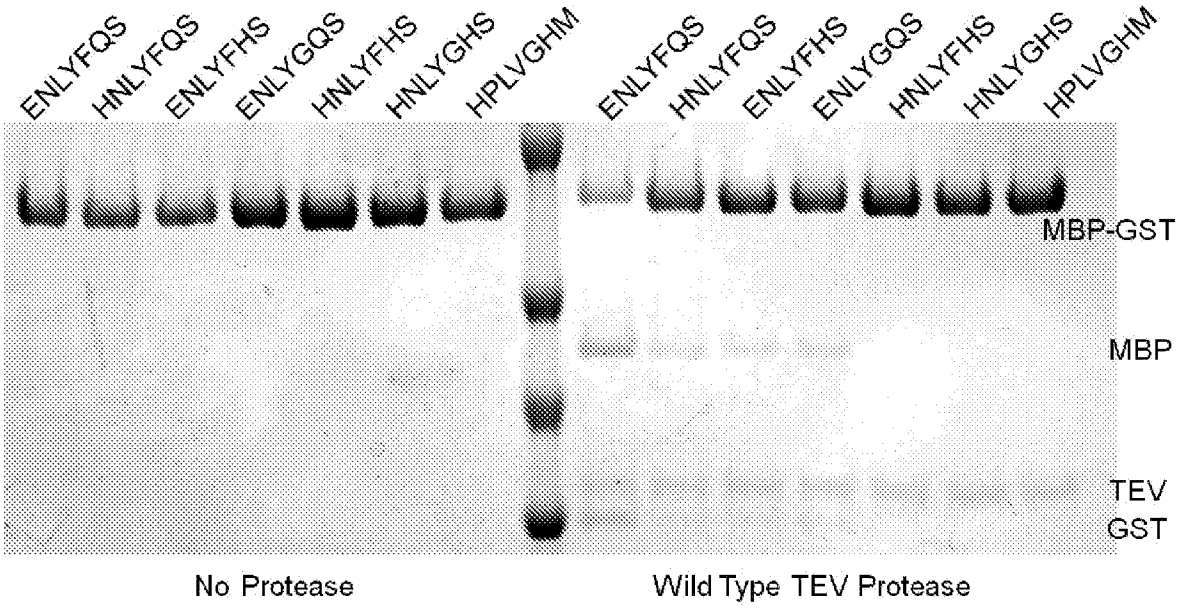

ENLYFQS
HNLYFQS
ENLYFHS
ENLYGQS
HNLYFHS
HNLYGHS
HPLVGHM
ENLYFQS
HNLYFQS
ENLYFHS
ENLYGQS
HNLYFHS
HNLYGHS
HPLVGHM
MBP-GST
MBP
TEV
GST
No Protease
Wild Type TEV Protease

FIG. 20

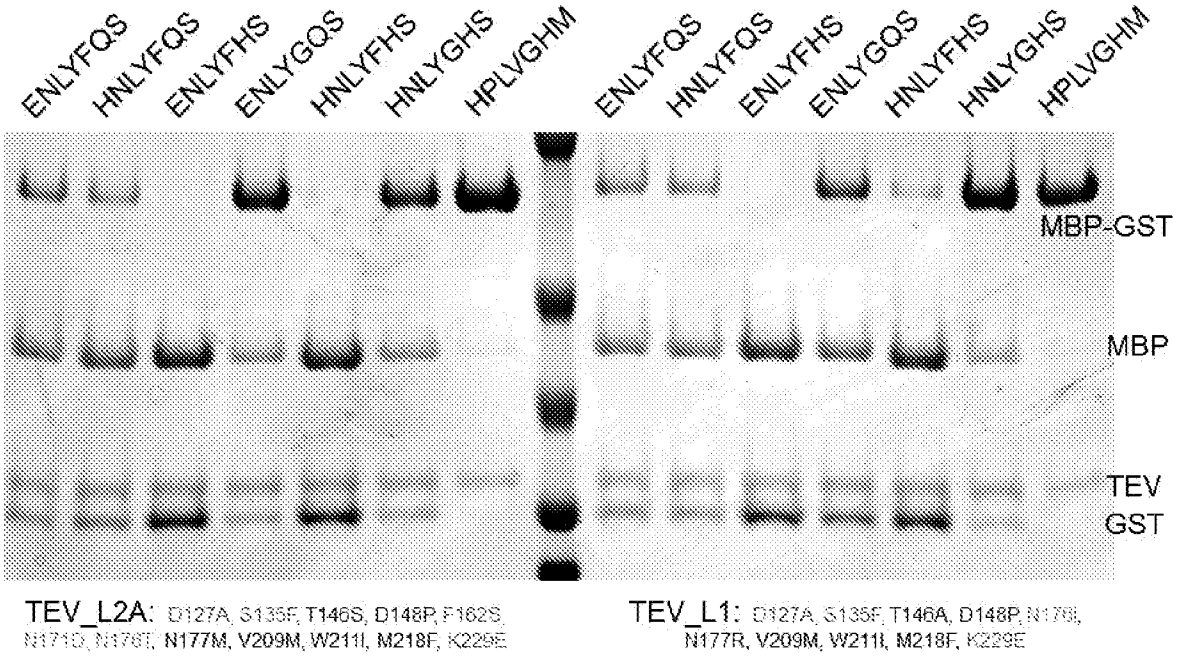
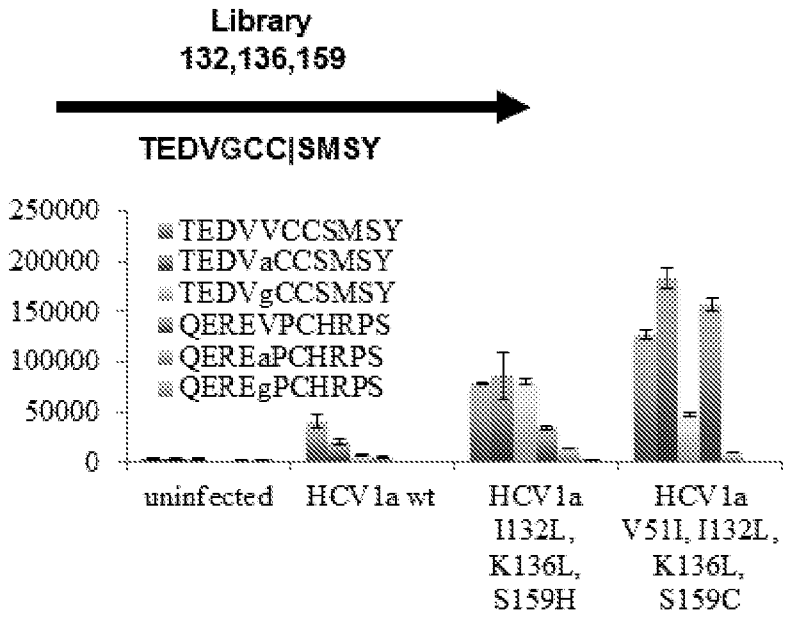

ENLYFQS
HNLYFQS
ENLYFHS
ENLYGQS
HNLYFHS
HNLYGHS
HPLVGHM
ENLYFQS
HNLYFQS
ENLYFHS
ENLYGQS
HNLYFHS
HNLYGHS
HPLVGHM
MBP-GST
MBP
TEV
GST
TEV_L2A: D127A, S135F, T146S, D148P, F162S, N171D, N176T, N177M, V209M, W211I, M218F, K229E
TEV_L1: D127A, S135F, T146A, D148P, N176I, N177R, V209M, W211I, M218F, K229E

FIG. 21


Library
132,136,159
TEDVGCC|SMSY
250000
200000
150000
100000
50000
0
TEDVVCCSMSY
TEDVaCCSMSY
TEDVgCCSMSY
QEREVPCHRPS
QEREaPCHRPS
QEREgPCHRPS
uninfected
HCV1a wt
HCV1a I132L, K136L, S159H
HCV1a V51I, I132L, K136L, S159C

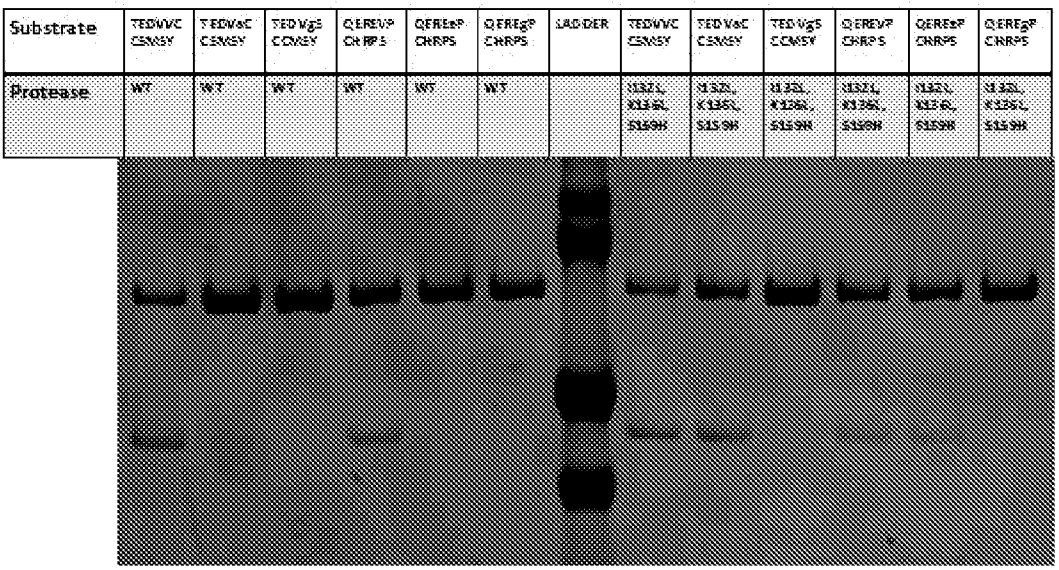
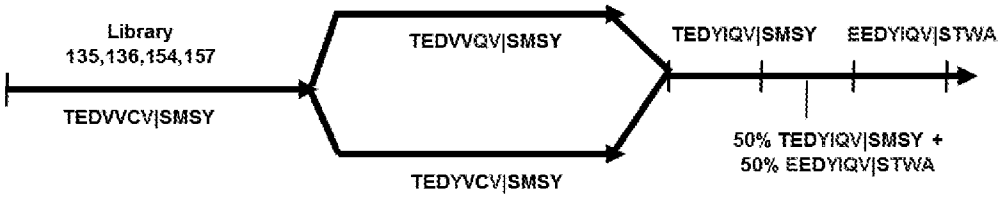

Substrate
Protease
WT
LADDER

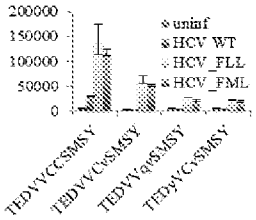
FIG. 22

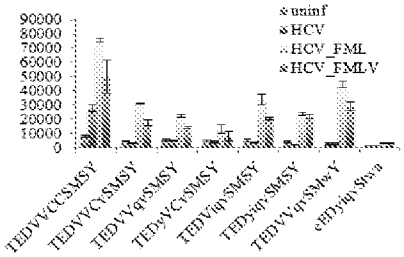


L135F	K136M	F154L
L135F	K136L	F154L
L135F	K136M	F154L
L135F	K136M	F154L
L135F	K136M	F154L
L135F	K136M	F154L
L135F	K136M	F154L
L135F	K136M	F154L

A5V					L135F	K136M	F154L	A157V	
A5V					L135F	K136M	F154L	A157V	D168Y
A5V		T19A			L135F	K136M	F154L	A157V	
					L135F	K136M	F154L	A157V	
A5V					L135F	K136M	F154L	A157V	
A5V					L135F	K136M	F154L	A157V	
A5V					L135F	K136M	F154L	A157V	
A5V					L135F	K136M	F154L	A157V	
			T72A	V78A	L135F	K136M	F154L	A157V	
					L135F	K136M	F154L	A157V	
					L135F	K136M	F154L	A157V	
					L135F	K136M	F154L	A157V	
					L135F	K136M	F154L	A157V	
			T72A	V78A	L135F	K136M	F154L	A157V	
A5V	E13K				L135F	K136M	F154L	A157V	
A5V					L135F	K136M	F154L	A157V	

		L135F	K136M	F154L	A156G	A157V	V158F	stop183G	stop184
		L135F	K136M	F154L	A156G	A157V	V158F	stop183G	stop184
866F		L135F	K136M	F154L	A156G	A157V	V158F	stop183G	stop184
		L135F	K136M	F154L	A156G	A157V	V158F	stop183G	stop184
		L135F	K136M	F154L	A156G	A157V	V158F	stop183G	stop184
		L135F	K136M	F154L	A156G	A157V	V158F	stop183G	stop184
		L135F	K136M	F154L	A156G	A157V	V158F	stop183G	stop184
		L135F	K136M	F154L	A156G	A157V	V158F	stop183G	stop184





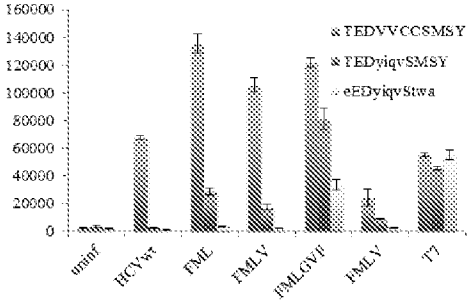


FIG. 23