Model System for High-Throughput Screening of Novel Human Immunodeficiency Virus Protease Inhibitors in *Escherichia coli*

Ting-Jen Cheng,¹ Ashraf Brik,² Chi-Huey Wong,² and Chen-Chen Kan¹*

Keck Graduate Institute of Applied Life Sciences, Claremont, California 91711,¹ and Department of Chemistry and the Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California 92037²

Received 10 February 2004/Returned for modification 28 February 2004/Accepted 5 March 2004

Novel human immunodeficiency virus (HIV) protease inhibitors are urgently needed for combating the drug-resistance problem in the fight against AIDS. To facilitate lead discovery of HIV protease inhibitors, we have developed a safe, convenient, and cost-effective *Escherichia coli*-based assay system. This *E. coli*-based system involves coexpression of an engineered β-galactosidase as an HIV protease substrate and the HIV protease precursor comprising the transframe region and the protease domain. Autoprocessing of the HIV protease precursor releases the mature HIV protease. Subsequently, the HIV protease cleaves β-galactosidase, resulting in a loss of the β-galactosidase activity, which can be detected in high-throughput screens. Using Food and Drug Administration-approved HIV protease inhibitors, this *E. coli*-based system is validated as a surrogate screening system for identifying inhibitors that not only possess inhibitory activity against HIV protease but also have solubility and permeability for in vivo activity. The usefulness of the *E. coli*-based system was demonstrated with the identification of a novel HIV protease inhibitor from a library of compounds that were prepared by an amide-forming reaction with transition-state analog cores. A novel inhibitor with a sulfonamide core of amprenavir, E2, has shown good correlation with the in vitro enzymatic assay and in vivo *E. coli*-based system. This system can also be used to generate drug resistance profiles that could be used to suggest therapeutic uses of HIV protease inhibitors to treat the drug-resistant HIV strains. This simple yet efficient *E. coli* system not only represents a screening platform for high-throughput identification of leads targeting the HIV proteases but also can be adapted to all other classes of proteases.

Human immunodeficiency virus (HIV) protease, essential for processing HIV viral polyproteins into individual structural proteins and replication enzymes during virus maturation, has been an attractive target for antiviral therapy development (9). Several HIV protease inhibitors designed to bind the active site of HIV protease have been shown to have in vivo efficacy and are currently in clinical use. This line of therapy, by itself or in combination with reverse transcriptase inhibitors, has revolutionized antiviral treatments and dramatically lowered the number of deaths due to AIDS (32). However, widespread use of HIV protease inhibitors has caused the rapid emergence of drug-resistant HIV proteases (8), rendering AIDS with no definitive cure. In these HIV proteases, mutations have been found in 49 of the 99 amino acids of the coding sequence. Substitutions at 18 or more positions are directly correlated with loss of responsiveness to HIV protease inhibitor treatment (26). The native TF comprises an 8-amino-acid transframe peptide and a 48-amino-acid p6pol protein. Purified recombinant p6pol protein was shown to have inhibitory effects on HIV protease activity (25). The intact HIV protease precursor, including TF and the protease domain, has a very low dimer stability relative to that of the mature enzyme and exhibits negligible levels of stable tertiary structure (33). Louis et al. (19) proposed that TF interacts with the dimer interface and thus destabilizes the dimeric structure to regulate HIV protease activity. The solution structure of the HIV protease precursor confirmed that TF hinders dimerization of the HIV protease domain and influences the catalytic activity of HIV protease (11).

The molecular mechanism of HIV protease activation is not completely clear yet. It has been suggested that the transframe region (TF), flanking the N terminus of the protease domain in the Pol gene, is involved in the regulation of HIV protease activity (26). The native TF comprises an 8-amino-acid transframe peptide and a 48-amino-acid p6pol protein. Purified recombinant p6pol protein was shown to have inhibitory effects on HIV protease activity (25). The intact HIV protease precursor, including TF and the protease domain, has a very low dimer stability relative to that of the mature enzyme and exhibits negligible levels of stable tertiary structure (33). Louis et al. (19) proposed that TF interacts with the dimer interface and thus destabilizes the dimeric structure to regulate HIV protease activity. The solution structure of the HIV protease precursor confirmed that TF hinders dimerization of the HIV protease domain and influences the catalytic activity of HIV protease (11).

Standard screening methods of HIV protease inhibitors include both in vitro and in vivo assays. While the in vitro enzyme kinetic assays use purified recombinant HIV proteases and noncovalently associated monomers. The active site of the enzyme is formed at the dimer interface and contains two conserved catalytic aspartic acid residues, one from each monomer (5). Instead of being translated as an active enzyme, HIV protease is synthesized as an integral part of the viral Gag-Pol polyprotein and needs to be activated to gain catalytic activity. Activation of the HIV protease requires appropriate folding and dimerization of the protease domain in the viral Gag-Pol polypeptide to form an active site, which then can catalyze the hydrolysis of peptide bonds at cleavage sites to release the mature HIV protease (34).

The molecular mechanism of HIV protease activation is not completely clear yet. It has been suggested that the transframe region (TF), flanking the N terminus of the protease domain in the Pol gene, is involved in the regulation of HIV protease activity (26). The native TF comprises an 8-amino-acid transframe peptide and a 48-amino-acid p6pol protein. Purified recombinant p6pol protein was shown to have inhibitory effects on HIV protease activity (25). The intact HIV protease precursor, including TF and the protease domain, has a very low dimer stability relative to that of the mature enzyme and exhibits negligible levels of stable tertiary structure (33). Louis et al. (19) proposed that TF interacts with the dimer interface and thus destabilizes the dimeric structure to regulate HIV protease activity. The solution structure of the HIV protease precursor confirmed that TF hinders dimerization of the HIV protease domain and influences the catalytic activity of HIV protease (11).

Standard screening methods of HIV protease inhibitors include both in vitro and in vivo assays. While the in vitro enzyme kinetic assays use purified recombinant HIV proteases and noncovalently associated monomers. The active site of the enzyme is formed at the dimer interface and contains two conserved catalytic aspartic acid residues, one from each monomer (5). Instead of being translated as an active enzyme, HIV protease is synthesized as an integral part of the viral Gag-Pol polyprotein and needs to be activated to gain catalytic activity. Activation of the HIV protease requires appropriate folding and dimerization of the protease domain in the viral Gag-Pol polypeptide to form an active site, which then can catalyze the hydrolysis of peptide bonds at cleavage sites to release the mature HIV protease (34).

The molecular mechanism of HIV protease activation is not completely clear yet. It has been suggested that the transframe region (TF), flanking the N terminus of the protease domain in the Pol gene, is involved in the regulation of HIV protease activity (26). The native TF comprises an 8-amino-acid transframe peptide and a 48-amino-acid p6pol protein. Purified recombinant p6pol protein was shown to have inhibitory effects on HIV protease activity (25). The intact HIV protease precursor, including TF and the protease domain, has a very low dimer stability relative to that of the mature enzyme and exhibits negligible levels of stable tertiary structure (33). Louis et al. (19) proposed that TF interacts with the dimer interface and thus destabilizes the dimeric structure to regulate HIV protease activity. The solution structure of the HIV protease precursor confirmed that TF hinders dimerization of the HIV protease domain and influences the catalytic activity of HIV protease (11).

Standard screening methods of HIV protease inhibitors include both in vitro and in vivo assays. While the in vitro enzyme kinetic assays use purified recombinant HIV proteases and noncovalently associated monomers. The active site of the enzyme is formed at the dimer interface and contains two conserved catalytic aspartic acid residues, one from each monomer (5). Instead of being translated as an active enzyme, HIV protease is synthesized as an integral part of the viral Gag-Pol polyprotein and needs to be activated to gain catalytic activity. Activation of the HIV protease requires appropriate folding and dimerization of the protease domain in the viral Gag-Pol polypeptide to form an active site, which then can catalyze the hydrolysis of peptide bonds at cleavage sites to release the mature HIV protease (34).

The molecular mechanism of HIV protease activation is not completely clear yet. It has been suggested that the transframe region (TF), flanking the N terminus of the protease domain in the Pol gene, is involved in the regulation of HIV protease activity (26). The native TF comprises an 8-amino-acid transframe peptide and a 48-amino-acid p6pol protein. Purified recombinant p6pol protein was shown to have inhibitory effects on HIV protease activity (25). The intact HIV protease precursor, including TF and the protease domain, has a very low dimer stability relative to that of the mature enzyme and exhibits negligible levels of stable tertiary structure (33). Louis et al. (19) proposed that TF interacts with the dimer interface and thus destabilizes the dimeric structure to regulate HIV protease activity. The solution structure of the HIV protease precursor confirmed that TF hinders dimerization of the HIV protease domain and influences the catalytic activity of HIV protease (11).
specific substrates, the in vivo assays use mammalian cells and the HIV viruses to evaluate the permeability as well as the in vivo antiviral activities of the potential drug leads. The complexity and time-consuming nature of the current protocols prompted us to explore the use of an Escherichia coli-based system as a simple and safe alternative approach for HIV protease inhibitor screening.

Several E. coli-based systems have been developed for in vivo HIV protease activity assays. Some systems have used reporter enzymes, like thymidylate synthase (16) or β-galactosidase (2), while other systems used the proteins involved in regulatory pathways, such as the lytic/lysogenic cycle of λ phage (22, 29), transcriptional control (6, 12, 30), or the cyclic AMP signaling pathway (7). These systems have limited success, however, in adaptation for high-throughput screening or in the detection of in vivo activity of HIV protease inhibitors.

We report here the development of an E. coli-based system in a high-throughput format and with the sensitivity to detect in vivo activity of HIV protease inhibitors. This system involves an engineered β-galactosidase to report in vivo HIV protease activity, and the HIV protease precursor (TF-PR) that comprises TF and the protease domain from the Gag-Pol polyprotein. The autoproteolytic processing of the HIV protease precursor occurs in the E. coli cells and results in the release of the mature HIV protease. Subsequently, the mature HIV protease cleaves the β-galactosidase, causing a loss of β-galactosidase activity. β-Galactosidase activity, which can be further quantitated by the enzymatic assay with specific chromogenic substrates, is thus inversely related to HIV protease activity. The sensitive and robust enzymatic assay of β-galactosidase enables the detection of β-galactosidase activity by using either the purified enzyme, crude cell lysates (27), or whole cells (23).

The validation of this system was performed with Food and Drug Administration (FDA)-approved HIV protease inhibitors and a library of new inhibitors against both the wild-type HIV protease and the drug-resistant HIV proteases. The ranking of the inhibitory potency of FDA-approved HIV protease inhibitors against the wild-type HIV protease in this E. coli-based system was shown to correspond well with what was determined in standard antiviral assays. A novel HIV protease inhibitor, E2, was identified as a potential drug lead from a two-cistron approach (Fig. 1A). The first cistron encodes the HIV protease, a pair of cleavage cassettes, encoding the decapeptide corresponding to the p6/PR cleavage site (Val-Ser-Phe-Asn-Phe-Pro-Gln-Ile-Thr-Glu) and the protease domain from the Gag-Pol polyprotein, were synthesized and ligated with Saull-digested pBAD-TOPO LacZ (7). The construct was designated pBAD-GalPR.

Coexpression of the HIV protease and β-Galactose was carried out by using a two-cistron approach (Fig. 1A). The first cistron encodes the HIV protease or TF-PR. Right before the stop codon of the first cistron was a Shine-Dalgarno sequence of AGGAGGAA for ribosome binding, which was followed with a start codon for translation of the second cistron, β-GalPR. The coding sequence for the HIV protease gene was amplified by PCR from HIV-1 NL4.3 DNA with oligonucleotide PR-F (ATACATGCGCTTCTGGTCG) as the forward primer and oligonucleotide PR-R (AGCCCACTGTTTGGTACG) as the reverse primer. The amplified fragments were then cloned into the unique Ncol restriction site of pCgl-βGalPM. The resulting plasmid was designated pPRWT or pTF-PRWT, respectively. E. coli DH5α was used as the host for plasmid preparations and for recombinant protein expression (27).

In vitro PCR-mediated mutagenesis. The HIV protease mutant with a mutation at residue 25 was constructed by PCR-mediated mutagenesis (1). The PCR-mediated mutagenesis includes three PCRs. The first PCR used pTF-PRWT as the template to generate the upstream fragment of the mutation site with TF-PR as the forward primer and PRD25N-R (ATTAAAGAAGCTTCAC) as the reverse primer. The resulting PCR product was sequenced by automatic DNA sequencing (DAVIS Sequencing, Davis, Calif.) and cloned into the unique Ncol restriction site of pCgl-βGalPM. The resulting construct was designated pPRD25NPR. The resulting construct was designated pPRD25N or pTF-PRD25N, respectively.

The HIV protease precursor variant containing the D30N mutation or the 64V mutation was similarly constructed by PCR-mediated mutagenesis as described previously (1). The oligonucleotides TF-PR and PRD30N-R (CTGGCGAATTATTTCTCTTTGGCAGC) or PR816V-R (TACTTGGGTTTCTTTAGAAGGTAATAC) were used as primers for the first-round PCR to generate the upstream fragment of the mutation site. The oligonucleotides PRD30N-F (AACACAGTTAGAAGAAATGTGTAAGTCC) and PR816V-F (GTTGACAATTTTAGCCTACTAATCTC) were used as primers for the second-round PCR to generate the downstream fragment containing the mutation site. The corresponding fragment pairs were then combined, and the final product was amplified by PCR to produce the final product, i.e., TF-PRD30N or TF-PR816V. When, the expression plasmid was digested with Ncol and recloned into pCgl-βGalPM. The resulting plasmid was designated pTF-PRD30N or pTF-PR816V. The sequences of all constructs were confirmed by automated DNA sequencing (DAVIS Sequencing, Davis, Calif.)

Diversity-oriented chemical synthesis of HIV protease inhibitors. Fifty micro-liters of each different carbonyl acid (3 μmol) was added to each well of a
reaction mixture, 50% (HBTU, 3/1 CHCl3/MeOH ratio; on the disappearance of the free amine monitored by thin-layer chromatography and kept at room temperature. The reactions went to completion in 1 h, based dimethyl formamide (DMF) was added. All of the reaction mixtures were mixed and washed with 1 N HCl, saturated aqueous NaHCO3, and brine, dried over the addition of brine, and extracted with ethyl acetate. The organic layer was contained the lysate from 0.3 OD630 equivalents of cells, except the lane for PR D25N, which contained 0.1 OD630 equivalents of cells expressing 3-hydroxy benzoic acid (102 mg, 0.67 mmol) in 6 ml of dry DMF was added to HBTU (253 mg, 0.67 mmol) followed by DIEA (230 mol of amine core 1 (Fig. 2) dissolved in anhydrous dimethyl formamide (DMF) was added. All of the reaction mixtures were mixed and kept at room temperature. The reactions went to completion in 1 h, based on the disappearance of the free amine monitored by thin-layer chromatography (10:1 CHCl3/Methanol ratio; Rf = 0.28) and analysis of the crude reaction mixture by electrospray ionization mass spectrometry (3).

Synthesis and purification of inhibitor E2. Core 1 was prepared as previously reported (31). To a solution of a free amine 1 (200 mg, 0.45 mmol), 2-methyl-3-hydroxy benzoic acid (102 mg, 0.67 mmol) in 6 ml of dry DMF was added to HBTU (253 mg, 0.67 mmol) followed by DIEA (230, 1.3 mmol) at 20°C in an argon atmosphere. The reaction mixture was stirred for 3 h, then quenched by the addition of brine, and extracted with ethyl acetate. The organic layer was washed with 1 N HCl, saturated aqueous NaHCO3, and brine, dried over MgSO4, filtered, and concentrated in vacuo. The crude product was purified by flash chromatography (CHCl3/Methanol), and the desired product was obtained in a 90% yield. The chemical structure of E2 was determined by nuclear magnetic resonance and electrospray ionization mass spectrometry (3).

Determination of in vitro IC50 of inhibitors against HIV protease. Recombinant HIV protease for enzymatic assays was prepared as described previously (3). The determination of kinetic parameters of HIV protease was performed at 37°C and 5.6 by using an F-2000 fluorescence spectrophotometer (Hitachi) and a Packard fluorescence spectrophotometer (Fusion-Universal microplate analyzer) for the microplate assay. For HIV protease, the Ks and Vs values for the fluorogenic peptide substrate 2-aminobenzoyl (Aba)-Thr-Ile-Nle-Phe-(p-NO2)-Gln-Arg-NH2 were determined by measuring the initial rate of hydrolysis at different substrate concentrations (2.5, 5.0, 10, 25, 50, and 100 μM) by monitoring the changes in fluorescence at a 420-nm emission and fitting the obtained data to the Michaelis-Menten equation with the GRAFIT program (version 3.0; Erithacus Software, Surrey, United Kingdom). Assays were run in 0.1 M morpholinethanesulfonic acid (MES) buffer containing 0.2 M NaCl and 1 mM dithiothreitol in a final volume of 200 μl in the wells of microplates. The enzyme concentration (30 μg/ml) that gave an ideal progress curve was used for the assays. The in vitro IC50 of the concentration required for 50% inhibition of enzyme activity, was determined from the percentage of inhibition rendered by E2 at various concentrations: 0.1, 0.5, 2.0, 4.0, 8.0, 12, and 16 nM. The Ki of the compound E2 was derived from the in vitro IC50 by using the formula for competitive inhibition: Ks = IC50/1+ [S]Ks.

Treatment of E. coli cells with HIV protease inhibitors. Pepstatin A was dissolved in water at 100 mM. Stock solutions of the FDA-approved HIV protease inhibitors were prepared by solubilizing the drugs in 100% dimethyl sulfoxide (DMSO) at 100 mM. A single colony of the cells bearing the respective plasmid, namely pTF-PRWT or pTF-PRmutant, was inoculated to 2 ml of Luria-Bertani medium with 100 μg of ampicillin/ml and incubated at 37°C for 2 h until the optical density at 630 nm (OD630) of the cells reached 0.6. The cells were then induced for protein expression with 0.2% arabinose and simultaneously treated with either HIV protease inhibitors at the indicated concentrations dissolved in 2% DMSO or 2% DMSO alone as the negative control. After incubation for an additional 3 h at 30°C, the cells were collected for Western blot analysis of proteins.

FIG. 1. Coexpression of HIV protease and β-galactosidase in E. coli. (A) The two-cistron constructs used in this study. The first cistron is the HIV protease precursor, which consists of the TF, including the transframe peptide (TFP) and p6pol, and the protease domain (PRWT). The second cistron is the engineered β-galactosidase, containing an insertion of the protease cleavage site, Val-Ser-Asn-Phe-Pro-Gln-Ile-Thr-Leu, at amino acid 131 (β-GalPR). In the pTF-PRD25N construct, the inactive HIV protease with the catalytic Asp25-to-Asn mutation was used. The arrows represent the cleavage sites of HIV protease. (B) Autoprocessing of the HIV protease precursor detected with HIV-1 protease antiserum. Cells bearing individual plasmids (indicated at the top of the blots) were induced with 0.2% arabinose for 30°C. The cells were then collected and processed for separation on 4 to 12% Bis-Tris gels followed by immunoblot analysis. M denotes the molecular weight standards. Each lane contained the lysate from 0.3 OD630 equivalents of cells, except the lane for PR D25N, which contained 0.1 OD630 equivalents of cells expressing PRD25N. (C) Cleavage of β-GalPR detected with antibodies against β-galactosidase. Lanes 1, 3, and 4 were the same as mentioned earlier, except that lane 2 contained the lysates from the cells expressing only β-GalPR. Each lane contained the lysate from 0.3 OD630 equivalents of cells.
Western blot analysis of proteins. After the cells were induced and treated as described above, they were collected by centrifugation and solubilized in Laemmli sample buffer (17) at a concentration of 0.01 OD630 cells per μl of sample buffer. The cellular proteins were separated on 4 to 12% Novex Bis-Tris polyacrylamide gels and electrotransferred onto polyvinylidene fluoride membranes for analysis with specific antibodies. The membranes were blocked with 5% nonfat milk in TTBS (10 mM Tris-HCl [pH 7.4], 500 mM NaCl, 0.1% Tween 20) for 1 h and then incubated with 1:5,000 anti-β-galactosidase monoclonal antibodies or 1:5,000 HIV-1 protease antiserum in 1% nonfat milk in TTBS for another hour. The bound antibodies were detected by chemiluminescence by horseradish peroxidase and Lumi-Glo chemiluminescence reagent.

Cell-based β-galactosidase activity assay. E. coli cells bearing each of the expression plasmids were grown as described, except that after incubation at 37°C for 2 h to an OD630 of 0.6, the cells were transferred in triplicate sets to the wells of 96-well microplates. Similarly, the cells were subjected to simultaneous induction for recombinant protein expression as well as drug treatment with HIV protease inhibitors at the indicated concentrations in 2% DMSO. For the screening of the compound library generated by the amide-forming reaction, DMF was used in place of DMSO. After incubation for an additional 3 h at 30°C, the cells were made permeable for the entry of the β-galactosidase substrate, i.e., ortho-nitrophenyl-β-D-galactopyranoside (ONPG), by diluting the cells in assay buffer (10 mM NaP [pH 7.3], 10 mM NaCl, 1 mM MgSO4, 5 mM β-mercaptoethanol) that contained 50 μg of polymyxin B sulfate/ml and 2% Triton X-100 (28). The cells were then incubated for 5 min at room temperature in 96-well microplates. After the OD630 of the cells were determined by using an Ultramark microplate imaging system (Bio-Rad, Hercules, Calif.), ONPG was added at a final concentration of 200 μM to start the enzymatic reactions at 37°C. Changes in the OD415, indicating the production of ortho-nitrophenol, were continuously monitored with the Ultramark microplate imaging system. The linear portion of the progression curve was used to determine the initial velocity with MATLAB (The MathWorks Inc., Natick, Mass.). The difference in OD415 (ΔA) was converted to the increase in product concentration ([C]) by using the equation ΔA = ε · ΔC · L, where L is the light path of 0.4 cm and the extinction coefficient (ε) of ortho-nitrophenol was 3,500 M⁻¹ cm⁻¹ under the assay conditions. The β-ga-
lactosidase activity was determined as the number of micromoles of ortho-nitrophenol produced per minute at 37°C. The β-galactosidase activity was then normalized against the amount of total cellular proteins, which was estimated by assuming that an OD<sub>630</sub> unit corresponds to 1.4 × 10<sup>9</sup> cells and that every 10<sup>9</sup> cells yield approximately 150 μg of proteins (23).

**RESULTS**

**Validation of detection of HIV protease activity in vivo in a high-throughput format.** It was reported that the engineered β-galactosidase that contains an insertion of HIV protease cleavage site at amino acid 131 still retains its activity, however, loses its activity when cleaved (2). Therefore, β-Gal<sup>PR</sup>, i.e., β-galactosidase with an insertion of Val-Ser-Phe-Asn-Phe-Pro-Gln-Ile-Thr-Leu, corresponding to the sequence of the p6/PR junction in HIV Gag-Pol polyprotein, was used as the reporter of HIV protease activity in this E. coli-based model system.

Our model system uses the HIV protease precursor that comprises the TF, including the transframe peptide and p6<sup>pro</sup> protein, and the protease domain for coexpression together with β-Gal<sup>PR</sup> (pTF-PR<sup>WT</sup>) (Fig. 1A). Coexpression under the same promoter control was carried out by using a two-cistron expression plasmid, where the HIV protease precursor was used as the first cistron and β-Gal<sup>PR</sup> was used as the second cistron. A modified form of the HIV protease precursor that contained a D25N mutation at the catalytic residue Asp25 to abolish the proteolytic activity (14) was constructed by PCR-mediated mutagenesis and was used as a control (pTF-PR<sup>D25N</sup>) (Fig. 1A).

The expression of recombinant proteins in the E. coli cells bearing the expression plasmid, pTF-PR<sup>WT</sup> or pTF-PR<sup>D25N</sup>, was analyzed by Western blotting with HIV-1 protease antiserum (Fig. 1B) and with β-galactosidase antibodies (Fig. 1C). The immunoblots confirmed that both HIV protease and β-galactosidase could be translated from a single mRNA in this two-cistron system. Most of the wild-type HIV protease precursor was autoprocessed, and the mature HIV protease was released in the E. coli cells (Fig. 1B, lane 3). The mature HIV protease subsequently cleaves β-Gal<sup>PR</sup>, resulting in the appearance of an extra band with a lower molecular weight in the lysates from the cells expressing TF-PR<sup>WT</sup> and β-Gal<sup>PR</sup> (Fig. 1C).

The cleavage of the engineered β-galactosidase by HIV protease would cause a loss of β-galactosidase activity. In an effort to develop a high-throughput screening platform, we optimized the experimental procedures to quantitatively determine the β-galactosidase activity that requires minimal sample preparations, i.e., with no need for cell lysis and further protein purification steps. We used Triton X-100 and polymyxin B sulfate to increase the permeability of the β-galactosidase substrate (28) and then started the β-galactosidase reaction with the addition of ONPG. Using this protocol, we determined that uninduced E. coli contained no detectable β-galactosidase activity and the E. coli cells expressing only β-Gal<sup>PR</sup> gave a β-galactosidase activity of 0.191 ± 0.005 μmol/min/mg of total E. coli proteins. When β-Gal<sup>PR</sup> was coexpressed with HIV protease, β-galactosidase activity changed to 0.041 ± 0.003 and 0.178 ± 0.007 μmol/min/mg for E. coli producing active and inactive HIV protease, respectively. These data confirmed E. coli β-galactosidase as a valid reporter for in vivo HIV protease activity.

**Dose-dependent effects of the FDA-approved HIV protease inhibitors on the autoproteolytic processing of the HIV protease precursor.** The in vivo inhibitory effects of the HIV protease inhibitors were evaluated by a peptide-based protease inhibitor, pepstatin A, and the FDA-approved small-molecule HIV protease inhibitors APV, IDV, NFV, RTV, and SQV. Drug permeability was evaluated by Western blot analysis in which E. coli cells were treated with HIV protease inhibitors without the use of permeabilization agents. As shown in Fig. 3, all small-molecule HIV protease inhibitors, except pepstatin A, inhibited the autoprocessing of the HIV protease precursor in a dose-dependent fashion. The dose-dependent inhibition of autoprocessing could be detected with 1.6 μM HIV protease inhibitors and reached a limit with 200 μM HIV protease inhibitors. The cleavage at the TFP-p6<sup>cod</sup> site, resulting in the release of p6<sup>pro</sup>-PR, still occurred in the presence of 5 mM HIV protease inhibitors (Fig. 3). The concentration of IDV and RTV needed for 50% inhibition of the release of mature HIV protease was shown to be between 8 and 40 μM, and the
concentrations of APV, NFV, and SQV needed for 50% inhibition of the release of mature HIV protease was between 1.6 and 8 μM.

**Dose-dependent effects of the HIV protease inhibitor on β-galactosidase activity.** Since in vivo inhibitory effects of the FDA-approved HIV protease inhibitors in *E. coli* were confirmed by the inhibition of autoprocessing of the HIV protease precursor, we further investigated if we could detect the in vivo potency of the HIV protease inhibitors by an increase in the cleavage-induced loss of β-galactosidase activity. After drug treatment for 3 h, the cells were processed for detection of β-galactosidase activity in 96-well microplates as described. As shown in Fig. 4, β-galactosidase activity increased exponentially as the concentration of HIV protease inhibitors increased. Consistent with the inhibition of autoprocessing of the HIV protease precursor, the increase in β-galactosidase activity could be detected with as little as 1.6 μM HIV protease inhibitors and showed little differences between 200 μM and 400 μM HIV protease inhibitors, except with IDV. Additionally, with 400 μM HIV protease inhibitors, β-galactosidase activity was around 0.145 μmol/min/mg of total proteins, which was a little lower than the maximal β-galactosidase activity detected in the presence of inactive HIV protease. It is important that permeabilization reagents were used to allow the entry of ONPG for the in vivo β-galactosidase assay only after cells were harvested from drug treatment. Thus, it could not have had any effect on the permeability of HIV protease inhibitors. Indeed, pepstatin A with poor cell permeability did not show any in vivo inhibitory activity (data not shown).

The IC₅₀ of a given HIV protease inhibitor against the wild-type HIV protease in the *E. coli* system was determined by curve-fitting with MATLAB (Fig. 4 and Table 1). IC₅₀ refers to the concentration of the HIV protease inhibitor required to reach a 50% inhibition of the cleavage-induced loss of β-galactosidase activity. The β-galactosidase activity detected in *E. coli* coexpressing TF-PRWT and β-GalPR in the absence of...
HIV protease inhibitors was referred to as 0% inhibition. The maximal \( \beta \)-galactosidase activity observed in \( E. coli \) containing pTF-PR\( ^{D25N} \) was regarded as 100% inhibition, which could theoretically be achieved by the most potent HIV protease inhibitor. As shown in Table 1, the \( IC_{50} \) for each HIV protease inhibitor was in the micromolar range, many times higher than what was determined in standard antiviral assays. Nevertheless, this \( E. coli \)-based system successfully detected the in vivo inhibitory activities of HIV protease inhibitors.

### Dose-dependent effects of HIV protease inhibitor on a drug-resistant D30N HIV protease mutant.

The usefulness of this model system for drug-resistant HIV proteases was further validated with a D30N HIV protease mutant, a variant observed in HIV clinical isolates and specifically conferring drug resistance to NFV.

The HIV protease D30N mutant was created by PCR-mediated mutagenesis and was used to replace the wild-type HIV protease domain in the two-cistron expression construct for \( -\text{GalPR} \) to create the construct pTF-PR\( ^{D30N} \). After the \( E. coli \) cells containing the expression plasmid pTF-PR\( ^{D30N} \) were induced with recombinant protein expression, \( \beta \)-galactosidase activity was determined by using whole cells in crude extracts, in a system developed to detect the presence of the wild-type HIV protease. This indicated that the D30N HIV protease mutant has a lower catalytic efficiency upon cleaving the substrate, i.e., \( \beta \)-galactosidase, causing an increase in \( \beta \)-galactosidase activity compared with that determined in the presence of the wild-type HIV protease. Therefore, this system is suitable for detecting the HIV protease variants whose activities were different from the wild-type HIV protease activity.

The drug response of the D30N HIV protease to HIV protease inhibitors was investigated by using an experimental protocol similar to that used in the study of the wild-type HIV protease (Fig. 5). It is important to stress that, in this study, 0% inhibition was defined as the \( \beta \)-galactosidase activity detected in the \( E. coli \) cells coexpressing the D30N HIV protease and \( \beta \)-Gal\( ^{PR} \) in the absence of HIV protease inhibitors. The reference to 100% inhibition was still the same as mentioned earlier. As shown in Fig. 5, the sensitivities toward APV, IDV, RTV, and SQV for the D30N HIV protease were similar to those for the wild-type HIV protease. On the other hand, compared to 75% inhibition against the wild-type HIV protease, 400 \( \mu \)M NFV only showed 30% inhibition against the D30N HIV protease. The \( IC_{50} \) of NFV for the D30N HIV protease mutant was thus determined to be greater than 400 \( \mu \)M, at least ninefold higher than the \( IC_{50} \) determined for the wild-type HIV protease (Table 1).

### Preparation of a library of compounds by amide-forming reaction and in situ screening of in vitro HIV protease activity.

Starting from the sulfonamide isostere core 1, a compound library was prepared in microplates, and was screened in situ against the HIV protease. The library was initially screened at 100 nM, and those reaching 50% or greater inhibition were selected for a second screening at 10 nM. The two rounds of screening revealed the preference of aromatic residues at the P2 position, such as benzoic acid. More than thirty acids of ortho-, meta-, para-substituted benzoic acids (Fig. 2) were coupled to the core and screened at 5 nM, in which E2 emerged as the best inhibitor from this library (Table 2). Following these results, E2 was synthesized and purified. The in vitro IC\( _{50} \)\( ^{S} \) and \( K_{i} \) values of the purified E2 against the wild-type HIV protease and three other mutants were subsequently determined (Fig. 2).

### Detection of in vivo inhibitory activity of the compound library against the wild-type HIV protease.

The library of compounds was also tested for inhibitory potency of HIV protease activity in the \( E. coli \)-based model system. The initial screening was conducted at 200 \( \mu \)M. Consistent with in vitro enzymatic studies, E2 was identified as the most potent compound in the \( E. coli \)-based model system. As shown in Fig. 6A, E2 inhibited autoprocessing of the HIV protease precursor at 1.6 \( \mu \)M. After determining \( \beta \)-galactosidase activity in the cells treated with different concentrations of E2, the in vivo IC\( _{50} \) of E2 in the \( E. coli \)-based system was determined as 87.3 \( \pm \) 25.4 \( \mu \)M, which is comparable to the potency of current drugs against HIV protease (Fig. 6B and Table 1).
Detection of in vivo inhibitory activity of E2 against the I84V HIV protease. Since the core of inhibitor E2 is similar to that of APV, we were interested in investigating the drug response of E2 against the APV-resistant HIV protease mutants. One of the key signature substitutions for APV resistance resides at the amino acid residue I84 (10). The in vivo EC₉₀, the inhibitor concentration that inhibits 90% of virus-induced cytotoxicity, of APV against I84V HIV protease mutant is twofold higher than the EC₉₀ against the wild-type HIV protease (13). The in vitro enzymatic assay also showed a threefold increase of the Ki of APV against the I84V HIV protease mutant compared to that of the wild-type HIV protease (13).

Again, the I84V HIV protease mutant was created by PCR-mediated mutagenesis and used to replace the wild-type HIV protease domain in the two-cistron system to construct the plasmid pTF-PR¹⁸⁴V. After inducing E. coli cells bearing the expression plasmid pTF-PR¹⁸⁴V, β-galactosidase activity was determined as described. The β-galactosidase activity in the cells expressing TF-PR¹⁸⁴V and β-GalPR was 0.081 ± 0.001 μmol/min/mg of total proteins. Again, this was higher than the β-galactosidase activity (0.041 ± 0.003 μmol/min/mg of total proteins) detected in the presence of the wild-type HIV protease. This confirmed that the I84V HIV protease mutant has a lower catalytic efficiency than the wild-type HIV protease in this E. coli-based system.

The drug response of the I84V HIV protease mutant toward E2 was investigated by using similar experimental protocols, as described, and with APV as a control (Fig. 6C and D). In this study, 0% inhibition was defined as the β-galactosidase activity detected in the E. coli coexpressing the I84V HIV protease and β-GalPR without the treatment of HIV protease inhibitors. The IC₅₀ of E2 against I84V HIV protease was determined as 141.3 ± 50 μM while the IC₅₀ of APV was 280 ± 39 μM. Data obtained with E2 demonstrated that subtle changes of the structure at the P2 position could generate new HIV protease inhibitors with improved binding affinities toward drug-resistant HIV proteases.

The results indicated that this E. coli-based model system is useful in serving multiple purposes, including quick identification of drug leads against the protease target and establish-
HIV protease inhibitors have been developed into effective antiviral therapies for treating AIDS in the past decade. However, the effectiveness of the existing HIV protease inhibitors has been hampered by the emergence of the drug-resistant variants against various drugs.

**DISCUSSION**

HIV protease inhibitors have been developed into effective antiviral therapeutics for treating AIDS in the past decade. However, the effectiveness of the existing HIV protease inhibitors has been hampered by the emergence of the drug-resistant HIV proteases. HIV protease variants can escape inhibition by current HIV protease inhibitors and thus pose a great medical need for novel HIV protease inhibitors. We report here a simple yet efficient *E. coli*-based system for screening the HIV protease inhibitors that are not only potent but also permeable for crossing the cell membrane. This system uses β-galactosidase as the reporter to adapt high-throughput screening and was validated with the in vivo inhibitory effects of existing HIV protease inhibitors against both the wild-type and drug-resistant HIV proteases. The ranking of IC_{50}s of HIV protease inhibitors against the HIV proteases, the wild type and the drug-resistant variants, determined in our *E. coli*-based system corresponded well with published data (24). Furthermore, we demonstrated rapid identification of a novel HIV protease inhibitor from a library of compounds prepared by diversity-oriented synthesis in microtiter plates.

The usefulness of this *E. coli*-based system was demonstrated with the FDA-approved HIV protease inhibitors and pepstatin A, a peptide-based aspartic protease inhibitor, as a control for cell permeability. Concentrations of each HIV protease inhibitor in this *E. coli*-based system ranged from 1.6 μM to as high as 5 mM. The small-molecule HIV protease inhibitors exerted in vivo inhibitory effects on the autoprocessing of the HIV protease precursor at as low as 1.6 μM (NFV and SQV) or 8 μM (IDV and RTV) (Fig. 3), and these were three magnitudes greater than the in vitro IC_{50}s in the nanomolar range against the purified wild-type HIV protease. Pepstatin A has an in vitro IC_{50} of 0.7 μM against the purified HIV protease (15). Therefore, if it were permeable, then one would expect to detect an in vivo IC_{50} for pepstatin A of around 700 μM. In fact, pepstatin A, up to 5 mM, failed to exert any inhibitory effects in the *E. coli*-based system. These data indicated that pepstatin A has low cell permeability; thus, it fails to diffuse passively across the cell membrane to inhibit autoprocessing of the HIV protease precursor intracellularly. This result suggested that the *E. coli*-based system could be used as an alternative method for rapidly selecting compounds with in vivo activities which require both potent inhibitory activity and desirable permeability.

The in vivo inhibition of HIV protease precursor autoprocessing by HIV protease inhibitors reached a limit at 200 μM. Interestingly, the cleavage at the TFP-p6 pol site, resulting in the release of p6pol-PR, still occurred in the presence of 5 mM HIV protease inhibitors (Fig. 3). Consistent with our results, in vivo studies with HIV-infected human T lymphocytes found an accumulation of the Pol intermediates, p6pol-PR-RT-IN and p6-PR, in the presence of 1 μM SQV (18), which was many times higher than the nanomolar concentrations used in standard antiviral assays. Therefore, it was reasonable that we observed residual activity of autoproteolytic processing in the presence of 5 mM HIV protease inhibitor in the *E. coli*-based system.

The in vivo inhibitory effects of HIV protease inhibitors were quantitated by determining the IC_{50} of a given protease inhibitor by using an in vivo β-galactosidase activity assay. We compared the IC_{50}s determined in our study with other reported data sets: (i) EC_{90}s determined from antiviral assays with mammalian cells and (ii) IC_{50}s obtained from another *E. coli*-based screening system (Table 1). Obviously, in *E. coli*-based systems, the drug concentrations needed to detect the in vivo inhibitory activity of HIV protease inhibitors are three magnitudes higher than what were needed for antiviral testing with mammalian cells. However, the ranking of IC_{50}s determined in our system was similar to that determined with standard antiviral tests (24). The higher IC_{50}s obtained from the *E. coli* system might be mainly due to an additional barrier posed by the bacterial cell wall.

We demonstrated the feasibility of this system for drug resistance enzymes by using an HIV protease variant that has a D30N mutation in the HIV protease coding sequence. The D30N mutation is an active-site mutation and represents the major mutation that confers drug resistance specifically against NFV. The structural analysis of binary complex D30N HIV protease with NFV indicated altered interactions with hydrophobic P2 side chains, resulting in variable substrate specificity and lower catalytic efficiency (20). The k_{cat}/K_{m} of D30N HIV protease was determined to be 0.089 min^{-1} μM^{-1} compared

---

**TABLE 2. Percentage of in vitro HIV protease activity detected in the presence of 5 nM inhibitors from the compound library***

<table>
<thead>
<tr>
<th>R-COOH</th>
<th>HIV protease activity (%) with 5 nM inhibitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>75</td>
</tr>
<tr>
<td>B1</td>
<td>70</td>
</tr>
<tr>
<td>C1</td>
<td>75</td>
</tr>
<tr>
<td>D1</td>
<td>70</td>
</tr>
<tr>
<td>E1</td>
<td>68</td>
</tr>
<tr>
<td>F1</td>
<td>65</td>
</tr>
<tr>
<td>A2</td>
<td>35</td>
</tr>
<tr>
<td>B2</td>
<td>74</td>
</tr>
<tr>
<td>C2</td>
<td>80</td>
</tr>
<tr>
<td>D2</td>
<td>50</td>
</tr>
<tr>
<td>E2</td>
<td>0</td>
</tr>
<tr>
<td>F2</td>
<td>90</td>
</tr>
<tr>
<td>A3</td>
<td>5</td>
</tr>
<tr>
<td>B3</td>
<td>80</td>
</tr>
<tr>
<td>C3</td>
<td>96</td>
</tr>
<tr>
<td>D3</td>
<td>30</td>
</tr>
<tr>
<td>E3</td>
<td>60</td>
</tr>
<tr>
<td>F3</td>
<td>60</td>
</tr>
<tr>
<td>A4</td>
<td>90</td>
</tr>
<tr>
<td>B4</td>
<td>94</td>
</tr>
<tr>
<td>C4</td>
<td>99</td>
</tr>
<tr>
<td>D4</td>
<td>88</td>
</tr>
<tr>
<td>E4</td>
<td>93</td>
</tr>
<tr>
<td>F4</td>
<td>95</td>
</tr>
<tr>
<td>A5</td>
<td>90</td>
</tr>
<tr>
<td>B5</td>
<td>70</td>
</tr>
<tr>
<td>C5</td>
<td>97</td>
</tr>
<tr>
<td>D5</td>
<td>86</td>
</tr>
<tr>
<td>E5</td>
<td>79</td>
</tr>
<tr>
<td>F5</td>
<td>40</td>
</tr>
</tbody>
</table>

*All measurements were run in triplicate, and the reported values are the means of these measurements.
to 0.41 min$^{-1}$ μM$^{-1}$ for the wild-type HIV protease (21). This indicated that the D30N HIV protease has a catalytic efficiency fivefold lower than the wild-type HIV protease does. While using the β-galactosidase activity as the reporter, we detected a roughly twofold-lower activity for the D30N HIV protease compared to that for the wild-type enzyme. This system thus could be used to identify HIV protease variants that have different catalytic efficiencies than the wild-type HIV protease.

The drug response of the D30N HIV protease determined in our studies was similar to the drug resistance profile determined in standard antiviral assays (Table 1). In our HIV protease inhibitor studies with the E. coli-based model system, the drug-resistant D30N HIV protease showed low responsiveness specifically to NFV but not to other HIV protease inhibitors. The IC$_{50}$s of APV, IDV, RTV, and SQV are similar or a little lower against the D30N HIV protease versus the wild-type HIV protease. However, the IC$_{50}$ of NFV for the D30N HIV protease increased at least ninefold compared with that for the wild-type HIV protease. This result suggested that the most effective agent among the five tested in this study for treating patients infected with the HIV strain carrying a D30N mutation in its protease coding sequence may be SQV, followed by RTV. This E. coli-based system could potentially provide a simple alternative method for generating drug resistance profiles that could be used in turn to suggest therapeutic use of HIV protease inhibitors to treat drug-resistant HIV strains.

The amide-forming reaction has been successfully demonstrated as a rapid synthesis method for preparing diversity-oriented libraries against HIV protease for in situ screening (4). The method was used to prepare a library of compounds containing an APV-like core attached to various P2 residues, in which E2 emerged as a novel HIV protease inhibitor with potency against both the wild type and the I84V HIV proteases. The amino residue I84 resides in the S2 binding pocket, and its mutation occurs during most of the HIV protease inhibitor treatments. The inhibitor E2, containing the 2-methyl-3-hydroxyl-benzamide portion at the P2 residue rather than the tetrahydrofuran ring at the P2 residue of APV, showed higher potency against the I84V HIV protease mutant.

The amide-forming reaction could be used to combine different transition-state analog cores with various P2 or P3 residues to produce new inhibitors. Together with the E. coli-based screening platform, we should be able to perform rapid lead discovery of novel HIV protease inhibitors targeting the active site of HIV proteases. It is plausible that this E. coli-based screening method developed for the HIV protease inhibitor screening can be easily modified for inhibitor screening for other proteases, especially proteases from infectious agents.

FIG. 6. In vivo dose-dependent effects of E2 against HIV proteases in the E. coli-based screening system. (A) Inhibition of autoprocessing of the wild-type HIV protease precursor. The drug treatment of E. coli cells containing the expression construct, pTF-PR WT, and Western blot analysis with HIV protease antiserum was performed as described. Each lane contained 0.3 OD$_{630}$ equivalents of cells, except the lane for PRD25N, which contained 0.1 OD$_{630}$ equivalents of cells expressing PRD25N. (B) Inhibition of cleavage-induced loss of β-galactosidase activity in the presence of the wild-type HIV protease. The E2-treated cells were processed for β-galactosidase activity detection by using whole cells in the 96-well microplates as described. (C and D) In vivo dose-dependent effects of APV (C) and E2 (D) against the I84V HIV protease mutant. The in vivo activity of the HIV protease inhibitors was determined by the inhibition of the cleavage-induced loss of β-galactosidase activity. The cells coexpressing TF-PR I84V and β-GalPR were treated with HIV protease inhibitors at the indicated concentrations. β-Galactosidase activities (micromoles per minute per milligram of total protein) in the treated E. coli cells were determined, and the data were used to generate the dose-response curves by curve-fitting with MATLAB. Representative results from three separate experiments are shown, and each experiment had a triplicate set of each sample.
that can cause serious diseases with unmet medical needs. Such proteases include the lethal factor from Bacillus anthracis and viral proteases from poliovirus and other viruses. The standard drug discovery process involves laborious in vitro kinetic assays followed by dangerous in vivo virus-induced cytolysis assays in biosafety level 3 and 4 labs. This E. coli-based assay system could be used as an efficient, convenient, and safe surrogate system for rapid identification of promising protease inhibitors with minimal exposure of human lives to infectious agents.

ACKNOWLEDGMENTS

We thank David Galas for manuscript review and critical discussions.

This work was supported in part by the Keck Graduate Institute of Applied Life Sciences and the NIH (grant CA 99898 to C.-C.K.).

REFERENCES