Conversion of Thymidylate Synthase into an HIV Protease Substrate

Jean-Jacques Kupiec, Stéphane Hazebrouck, Thierry Leste-Lasserre, Pierre Sonigo

Conversion of Thymidylate Synthase into an HIV Protease Substrate*

Jean-Jacques Kupiec‡, Stephane Hazebrouck, Thierry Leste-Lasserre, and Pierre Sonigo‡

From the Génétique des Virus, ICGM-CNRS UPR 415, Institut Cochin de Génétique Moléculaire, 22 rue Méchain, 75014 Paris, France

Thymidylate synthase (TS) is an essential enzyme of DNA metabolism. We have carried out an extensive insertional mutagenesis of the *Escherichia coli* TS gene (*thyA*) using three different methods. Insertion of exogenous sequences at unique restriction sites or at random positions produced defective mutants, whereas comparison of TS sequences from different species allowed us to identify six zones permissive for insertions of exogenous sequences. The insertion of Human immunodeficiency virus type 1 (HIV-1) protease substrate sequences into the permissive sites converted TS to an HIV-1 protease substrate, and the *in vivo* cleavage of these insertions by the cloned HIV-1 protease conferred a thymidylate synthase-deficient phenotype in some of our *E. coli* mutant strains. In agreement with crystallographic data, these results show that the permissive sites are located in regions of the TS protein not essential for enzyme activity and accessible to cleavage by HIV protease. These results also show that it is possible to control a growth phenotype in *E. coli* through the protease-mediated destruction of an essential metabolic enzyme. Because both wild type and thymidylate synthase-deficient phenotypes are selectable on the appropriate growth medium, these *thyA* mutants could be used for genetic selections of protease inhibitors and analysis of protease specificities.

Cleavage of a precursor polypeptide by a protease is a general mechanism for the regulation of physiological processes (1–4). Human immunodeficiency virus (HIV-1)[1] provides an example. In retroviruses a viral protease is responsible for cleavage of the polyproteins encoded by the *gag* and *pol* genes, producing mature structural and enzymatic proteins. This process is essential for generating infectious virus particles (for review see Ref. 5). At the junctions of the protein domains, within the precursors, lie peptide target sequences that are recognized and cleaved by the virally encoded protease. Although some degree of similarity exists between these target sequences, no consensus sequence has been found to predict cleavability. Moreover, other sequences in cellular proteins have also been found to be cleaved by HIV-1 protease (Ref. 6 and references therein). In order to study the determinants of this substrate specificity, a system allowing the genetic selection of HIV protease activity is much needed. Therefore, we

decided to insert HIV-1 protease target sites into thymidylate synthase (TS), which is a selectable marker in *Escherichia coli*. This could make possible the genetic selection of protease mutants with altered specificities or protease inhibitors and substrates from large libraries of random peptides (see Fig. 1 for a detailed explanation). In fact, it has already been shown that HIV-1 Gag and Gag-Pol precursor polyproteins are correctly processed by HIV-1 protease when they are co-expressed in *E. coli* (7, 8) and that an HIV-1 protease target sequence could be inserted into *E. coli* β-galactosidase (9). In this latter case, the enzyme retained its activity, and when the cloned HIV-1 protease was coexpressed in the same strain, β-galactosidase was cleaved and inactivated. However, conditional growth has not been reported with this system. Similarly, HIV-1 protease, human rhinovirus protease 3C, or zucchini yellow mosaic virus protease target sequences have been inserted into *E. coli* proteins responsible for resistance to tetracycline or sensitivity to streptomycin (10–12). In these cases, coexpression of these mutants with their cognate protease impaired bacterial growth in the presence of the appropriate antibiotic. We thought that insertion of protease target sequences into TS could provide a valuable system because positive and negative selection media are available allowing the selection of both wild type and thymidylate synthase-deficient phenotypes (13, 14). These two types of selection should not only permit to search for inhibitors as in the case of previous systems based on antibiotic sensitivity phenotypes but also to study protease specificity (see Fig. 1 for an overview). TS is an essential enzyme of DNA metabolism catalyzing dTMP synthesis. Although TS has been characterized by various techniques, including x-ray crystallography (15, 16), insertional mutagenesis has not been done.

In order to obtain an *E. coli* strain with a thymidylate synthase-deficient phenotype conferred by HIV-1 protease activity, the first steps consist of mapping tolerant sites for insertion of exogenous sequences within the TS protein and to determine, among these sites, those that are accessible to the HIV-1 protease. We report here that we have proceeded to an extensive insertional mutagenesis of the *E. coli* TS gene (*thyA*) and found several regions permissive to insertions of HIV-1 protease target sequences. Further analysis showed that these insertions converted TS to an HIV-1 protease substrate and that the *in vivo* cleavage of these insertions conferred a thymidylate synthase-deficient phenotype in some of the *E. coli* mutated strains.

MATERIALS AND METHODS

Cloning, sequencing, site-directed mutagenesis, and immunoblotting were done according to standard procedures.

Plasmids

pTZthyA—The 1300-base pair *thyA* fragment was first subcloned from pBTAH2 (17) into the phagemid pTZ 18R (Pharmacia Biotech Inc.) using *Hin*dIII sites. The *Hin*dIII sites were destroyed and replaced by *Xho*I sites by site-directed mutagenesis.

[1] The abbreviations used are: HIV-1, human immunodeficiency virus type 1; TS, thymidylate synthase.

growth on

positive selection medium   negative selection medium

protease target sequence

essential protein (T.S.)

$+$ $-$

<u>inactive protease:</u> no cleavage, active T.S., wild type phenotype

essential protein (T.S.)

$-$ $+$

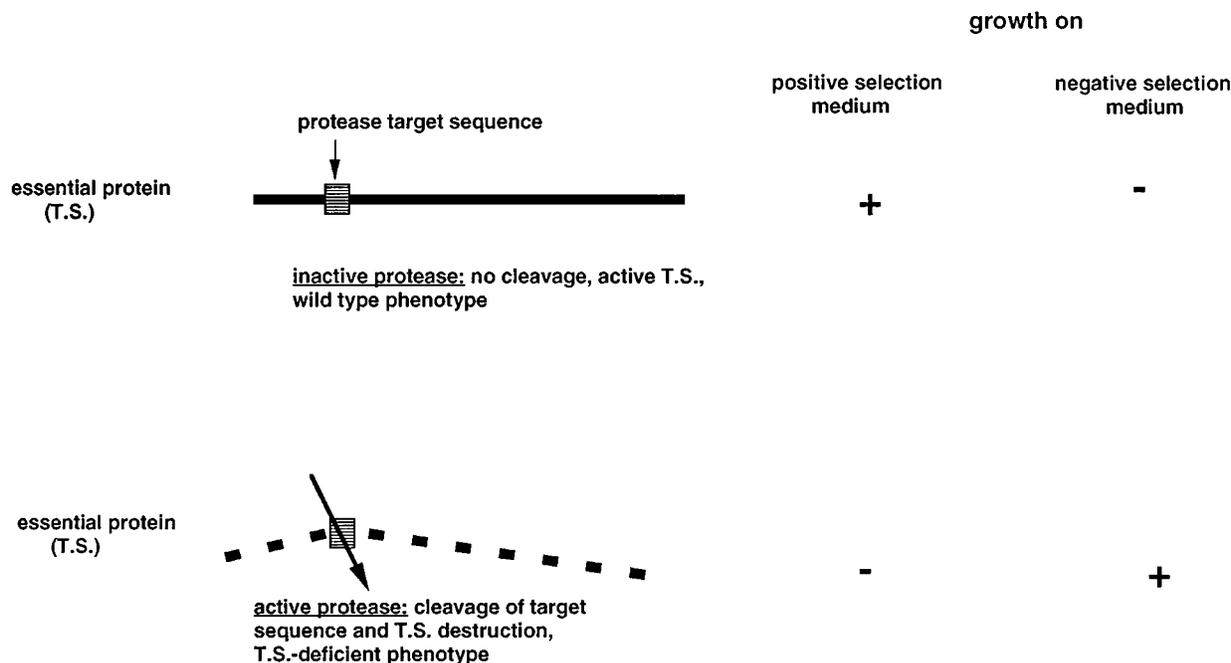<u>active protease:</u> cleavage of target sequence and T.S. destruction, T.S.-deficient phenotype

FIG. 1. **Genetic selection of protease inhibitors and substrates.** Insertion of a protease target sequence within an essential protein and expression of the cognate protease confers a protease-dependent phenotype. Insertion within TS allows selection of both wild type and thymidylate synthase-deficient phenotypes on positive and negative selection media, respectively (see "Materials and Methods" for a description of these media). These two types of selection allow a wide range of applications. For example: 1) Protease inhibitors could be directly selected on the positive selection medium from an endogenously expressed library of random peptides. The expression of the protease confers a thymidylate synthase-deficient phenotype, but those cells expressing a protease inhibitor will revert to a wild type phenotype and grow as colonies. 2) Protease substrates could also be selected from random sequences but this time inserted into TS itself. Those cells carrying a substrate sequence inserted into TS will have a thymidylate synthase-deficient phenotype and grow as colonies on the negative selection medium. By means of this method, exhaustive collections of protease substrate sequences could be collected, allowing the analysis of proteases specificity and the design of inhibitors. 3) Protease mutants could also be easily analyzed. For example, substrate specificity could be analyzed for protease mutants that are resistant to a currently known inhibitor, allowing the design of a new and specific inhibitor.

pSUthyA—*thyA* was also cloned into pSU 18 (18) by subcloning the *Xho*I insert from pTZ*thyA* into the *Sal*I site of pSU 18.

pTZprt⁺ and *pSUprt⁺*—The 2865-base pair HIV-1 *pol* gene fragment, corresponding to the protease and reverse transcriptase domains, was subcloned from pBRT1*prt⁺* (7) into pTZ 19 (Pharmacia) and pSU 19 (18), using the *Eco*RI and *Sal*I sites.

pTZprt⁻ and *pSUprt⁻*—The 1470-base pair HIV-1 *pol* fragment, corresponding solely to the reverse transcriptase domain, was also subcloned from pBRT3*prt⁻* (7) into pTZ 19 and pSU 19 using the *Eco*RI and *Sal*I sites. pBRT1prt⁺ and pBRT3*prt⁻* clones were kindly provided by the National Institutes of Health (AIDS Research and Reference Reagent Program).

HIV-1 Protease Target Sequences

Complementary oligonucleotides corresponding to HIV-1 protease target sequences were synthesized, hybridized, and cloned into *thyA*. The sequence for S1 is VSFNFPQITL (P6/protease junction of the HIV-1 Gag precursor); the DNA coding for this sequence has a *Hin*dIII site. The sequence for S2 is DRQGTVSFNFPQITLWQRPL (P6/protease junction of the HIV-1 Gag precursor). S3 is 50 amino acids from the P6/protease junction of the HIV-1 Gag precursor, encompassing S2. The sequence for S4 is IRKVLFLDGI (reverse-transcriptase/integrase junction of the HIV-1 Gag precursor).

Nomenclature of thyA Mutants

The three first letters designate the site of insertion (*i.e., ISE* means insertion at site E) and the two following characters designate the sequence inserted (*i.e., ISES1* means S1 inserted at site E).

Oligonucleotide-directed Mutagenesis

Mutagenesis was made according to Kunkel *et al.* (19). Oligonucleotides were synthesized to introduce an *Hpa*I site into *thyA* at sites A–G (Fig. 2). Annealing and extension reactions were performed on the single stranded form of the pTZ*thyA* phagemid DNA. Mutants were termed *ISA–ISG*, respectively.

Δ-thyA E. coli Strains, Selection Media, and Phenotypes

β-1083 is the HB101 Δ-*thyA*::Em. auxotroph for leucine, proline, and thymidine. It grows at 30 °C. β-1308 is the Δ-*thyA*::Em derivative of the wild type MG 1655 *E. coli* K12 (F⁻ λ⁻). Minimal medium contains citric acid (1.6 g/liter), MgSO₄ (0.4 g/liter), D-Mannitol (10 g/liter), NH₄Cl (2 g/liter), and K₂HPO₄ (17.4 g/liter). Müeller Hinton medium (Difco) is a rich medium devoid of thymidine. Positive selection medium is minimal or Müeller-Hinton medium. Negative selection medium contains minimal or Müeller-Hinton medium complemented with thymidine (100 μg/ml) and trimethoprim (20 μg/ml in minimal medium and 200 μg/ml in Müeller-Hinton medium). Nonselective medium is minimal or Müeller-Hinton medium complemented with thymidine (100 μg/ml). Wild type and thymidylate synthase-deficient phenotypes grow on positive and negative selection media, respectively. Mixed phenotypes grow on both media.

Antibodies and Protein Extracts for Immunoblotting Experiments

Antibody A is a polyclonal serum from a rabbit injected with the whole TS protein (kindly provided by Dr. F. Maley, Albany, NY). We also prepared antibody B, a polyclonal serum from a rabbit injected with a synthetic peptide corresponding to amino acids 241–264 of the *E. coli* TS protein. Protein extract preparation contained 1–1.5 ml of bacteria (OD = 0.6–1) were washed in ice-cold Tris-HCl, pH 7.4, resuspended in 50 μl of distilled water and mixed with 50 μl of 2 × Laemmli buffer. This suspension was then boiled for 5 min and centrifuged at 12,000 × *g* for 10 min, and the supernatant was stored at −80 °C.

Measurement of Thymidylate Synthase Activity

We measured tritium released from [³H]dUMP after the enzyme reaction was terminated, as described by Roberts (20). For protein extract preparation, pellets corresponding to 30-ml cultures of β-1083 bacteria transfected with one of the pTZ*thyA* mutants (OD = 1) were resuspended in 2 ml of buffer (0.04 mM phenylmethylsulfonyl fluoride, 4.8 mM 2-mercaptoethanol, 2 mM Tris-HCl, pH 7.5) and sonicated until all the cells were disrupted (4 × 30 s). Cellular debris was removed by centrifugation (5 min at 13,000 × *g*), and 50 μl of glycerol were added.

A

```
                                      19          20
Ecoli  ---------------------------MKQYLELMQKVLDEGTQKN----------DRTGTGTLSIFG-HQMRFNLQ-DGFPLVTTKRCHLRSIIHEL
Bsubt. --------------------------MKQYKDFCRHVLEHGEKKG----------DRTGTGTISTFG-YQMRFNLR-EGFPMLTTKKLHFKSIAHEL
TN4003 ------------------MYNPFDEAYHGLCEEILEIGNRRD----------DRTHTGTISKFG-HQLRFDLT-KGFPLLTTKKVSFKLVATEL
Lcasei -----------------MLEQPYLDLAKKVLDEGHFKP----------DRTHTGTYSIFG-HQMRFDLS-KGFPLLTTKKVPFGLIKSEL
Bphi3t -----------MTQFDKQYNSIIKDIINNGISDEEFDVRTKWDSDGTPAHTLSVMS-KQMRFDNS--EVPILTTKKVAWKTAIKEL
BT4    ------------------------MKQYQDLIKDIFENGYETD----------DRTGTGTIALFG-SKLRWDLT-KGFPAVTTKKLAWKACIAEL
Mouse  MLVVGSELQS-----DAQQL-SAEAPRHGELQYLRQVEHILRCGFKKE----------DRTGTGTLSVFG-MQARYSLR-DEFPLLTTKRVFWKGVLEEL
Rnorv. MLVEGSELQS-----GAQQP-RTEAPQHGELQYLRQVEHIMRCGFKKE----------DRTGTGTLSVFG-MQARYSLR-DEFPLLTTKRVFWKGVLEEL
Human  MPVAGSELPRRPLPPAAQERDAEPRPPHGELQYLGQIQHILRCGVRKD----------DRTGTGTLSVFG-MQARYSLR-DEFPLLTTKRVFWKGVLEEL
Athal. -------------HEEYLYLNLVKEIISNGNLKD----------DRTGTGTLSKFG-CQMKFNLR-RNFPLLTTKRVFWRGVVEEL
Scere. ---------------MTMDGKNKEEEQYLDLCKRIIDEGEFRP----------DRTGTGTLSLFAPPQLRFSLRDDTFPLLTTKKVFTRGIILEL
VZvir. -----------MGDLSCWTKVPGFTLTGELQYLKQVDDILRYGVRKR----------DRTGIGTLSLFG-MQARYNLR-NEFPLLTTKRVFWRAVVEEL
Pcari. --------------MVNAEEQQYLNLVQYIINHGEDRP----------DRTGTGTLSVFAPSPLKFSLRNKTFPLLTTKRVFIRGVIEEL
HVatel -----------MEELHAEHQYLSQVKHILNCGNFKH----------DRTGVGTLSVFG-MQSRYSLE-KDFPLLTTKRVFWRGVVEEL
Lmajor ----------CKYVPRNHEERQYLELIDRIMKTGIVKE----------DRTGVGTISLFG-AQMRFSLRDNRLPLLTTKRVFWRGVCEEL
Calb.  -----------------TVSPNTAEQAYLDLCKRIIDEGEHRP----------DRTGTGTKSLFAPPQLRFDLSNDTFPLLTTKKVFSKGIIHEL
Pchab. -------------------HPEYQYLNIIYDIIMHGNKQD----------DRTGVGVLSKFG-YMMKFNLS-EYFPLLTTKKLFVRGIIEEL
                         *               SITE A   *  *              * ***          **
```

```
          76        77        87                                                          88
Ecoli  LWFLQGDTN-IAYLHENN-------VTIWDEWADEN----------------------------------------------GDLGPVYGKQWRAW
Bsubt. LWFLKGDTN-VRYLQENG-------VRIWNEWADEN----------------------------------------------GELGPVYGSQWRSW
TN4003 LWFIKGDTN-IQYLLKYN-------NNIWNEWAFENYVQSDDYHGPDMTDFGHRSQQDPEFNEQYKEEMKKFKERILNDDAFAKKYGNLGNVYGKQWRDW
Lcasei LWFLHGDTN-IRFLLQHR-------NHIWDEWAFEKHVKSDEYHGPDMTDFGHRSQKDPEFAAVYHEEMAKFDDRVLHDDAFAAKYGDLGLVYGSQWRAW
Bphi3t LWIWQLKSNDVTELNKMG-------VHIWDQWKQED----------------------------------------------GTIGHAYGFQLGKK
BT4    IWFLSGSTN-VNDLRLIQHDSLIQGKTVWDENYENQAKDLG----------------------------------------------YHSGELGPIYGKQWRDF
Mouse  LWFIKGSTN-AKELSSKG-------VRIWDANGSRDFLDSLGFS----------------------------------------------ARQEGDLGPVYGFQWRHF
Rnorv. LWFIKGSTN-AKELSSKG-------VRIWDANGSRDFLDSLGFS----------------------------------------------ARQEGDLGPVYGFQWRHF
Human  LWFIKGSTN-AKELSSKG-------VKIWDANGSRDFLDSLGFS----------------------------------------------TREEEGDLGPVYGFQWRHF
Athal. LWFISGSTN-AKVLQEKG-------IRIWDGNASRAYLDGIGLT----------------------------------------------EREEGDLGPVYGFQWRHF
Scere. LWFLAGDTD-ANLLSEQG-------VKIWDGNGSREYLDKMGFK----------------------------------------------DRKVGDLGPVYGFQWRHF
VZvir. LWFIRGSTD-SKELAAKD-------IHIWDIYGSSKFLNRNGFH----------------------------------------------KRHTGDLGPIYGFQWRHF
Pcari. LWFIRGETD-SLKLREKN-------IHIWDANGSREYLDSIGLT----------------------------------------------KRQEGDLGPIYGFQWRHF
HVatel LWFIRGSTD-SKELAASG-------VHIWDANGSRSYLDKLGLF----------------------------------------------DREEGDLGPVYGFQWRHF
Lmajor LWFLRGETS-AQLLADKD-------IHIWDGNGSREFLDSRGLT----------------------------------------------ENKEMDLGPVYGFQWRHF
Calb.  LWFVAGSTD-AKILSEKG-------VKIWEGNGSREFLDKLGLT----------------------------------------------HRREGDLGPVYGFQWRHF
Pchab. LWFIRGETN-GNTLLEKN-------VRIWEANGTREFLDNRKLF----------------------------------------------HREVNDLGPIYGFQWRHF
           *           *    SITE B    *                              SITE C                 *  **  *
```

```
          104      105                             157                      188
Ecoli  PTP--------DGRHIDQITTVLNQLKNDPDSRRIIVSAWNVGELDKMALAPCHAFFQFYVADG-----------------KLSCQLYQRSCDVFLGLPF
Bsubt. RGA--------DGETIDQISRLIEDIKTNPNSRRLIVSAWNVGEIDKMALPPCHCLFQFYVSDG-----------------KLSCQLYQRSADVFLGVPF
TN4003 EDK--------NGNHYDQLKSVIQQIKTNPNSRRHIVSAWNVGEIDKMALPPCHTMFQFYVQEG-----------------KLNCQLYQRSADIFLGVPF
Lcasei HTS--------KGDTIDQLGDVIEQIKTHPYSRRLIVSAWNPEDVPTMALPPCHTLYQFYVNDG-----------------KLSLQLYQRSADIFLGVPF
Bphi3t NRS--------LNGEKVDQVDYLLHQLKNNPSSRRHITMLWNPDDLDAMALTPCVYETQWYVKQG-----------------KLHLEVRARSNDMALGNPF
BT4    -----------GGVDQIIEVIDRIKKLPNDRRQIVSAWNPAELKYMALPPCHMFYQFNVRNG-----------------YLDLQWYQRSVDVFLGLPF
Mouse  GAEYKDMDSDYSGQGVDQLQKVIDTIKTNPDDRRIIMCAWNPKDLPLMALPPCHALCQFYVVNG-----------------ELSCQLYQRSGDMGLGVPF
Rattus GADYKDMDSDYSGQGVDQLQKVIDTIKTNPDDRRIIMCAWNPKDLPLMALPPCHALCQFYVVNG-----------------ELSCQLYQRSGDMGLGVPF
Human  GAEYRDMESDYSGQGVDQLQRVIDTIKTNPDDRRIIMCAWNPRDLPLMALPPCHALCQFYVVNS-----------------ELSCQLYQRSGDMGLGVPF
Athal. GAKYTDMHADYTGQGFDQLLDVINKIKNNPDDRRIIMSAWNPSDLKLMALPPCHMFAQFYVANG-----------------ELSCQMYQRSADMGLGVPF
Scere. GAKYKTCDDDYTGQGIDQLGVIHKLKTNPYDRRIIMSAWNPADFDKMALPPCHIFSQFYVSFPKEGEG--------SGKPRLSCLLYQRSCDMGLGVPF
VZvir. GAEYKDCQSNYLQQGIDQLQTVIDTIKTNPESRRMIISSWNPKDIPLMVLPPCHTLCQFYVANG-----------------ELSCQVYQRSGDMGLGVPF
Pcari. GAEYIDCKTNYIGQGVDQLANIIQKIRTSPYDRRILLAMALEKMALPPCHMFCQFYVHIPSNNHR-----------PELSCQLYQRSCDMGLGVPF
HVatel GAEYQGLKHNYGGEGVDQLKQIINTIHTNPTDRRMLMCAWNVLDVPKMALPPCHVLSQFYVCDG-----------------KLSCQLYQRSADMGLGVPF
Lmajor GADYKGFEANYDGEGVDQIKLIVETIKTNPNDRRLLVTAWNPCALQQFYVNTDTS-----------------ELSCMLYQRSCDMGLGVPF
Calb.  GAEYKDCDSDYTGQGFDQLQDVIKKLKTNPYDRRIIMSAWNPPDFAKMALPPCHVFCQFYVNFPTSSPDPNNPKQAKTAKPKLSCLLYQRSCDMGLGVPF
Pchab. GAEYTDMHADYKDKGVDQLKNIINLIKNDPTCRRIILCAWNVKDLDQMALPPCHILCQFYVFDGK-----------------LSCIMYQRSCDLGLGVPF
       SITE D       **      *   **      *  *  ***  *      *   SITE E               **  *  **  **
```

```
                                                 235     236  237     242       243
Ecoli  NIASYALLVHMMAQQCDLEVGDFVWTGGDTHLYSNHMDQTHLQLSREPRPLPKLIIKRKP----ESIFDY--------RFEDFEIEGYDPHPGIKAPVAI
Bsubt. NIASYALLTMIIAHVTGLEPGEFIHTFGDVHIYQNHIEQVNLQLENRDVRPLPQLRFARKV----DSIFNF--------AFEDFIIEDYDPHPHIKGAVSV
TN4003 NIASYALLTHLVAKECGLEVGEFIHTFGDAHIYSNHMDAIHTQLSRDSYLPPQLKINTD-----KSIFDI--------NYEDLELINYESHPAIKAPIAV
Lcasei NIASYALLTHLVAHECGLEVGEFIHTFGDAHLYVNHLDQIKEQLSRTPRPAPTLQLNPDK-----HDIFDF--------DMKDIKLLNYDPYAIKAPVAV
Bphi3t NVFQYNVLQRMIAQVTGYELGEYIFNIGDCHVYTRHIDNLKIQMEREQFEAPELWINPEV----KDFYNF--------TVDDFKLINYKHGDKLLFEVAV
BT4    NIASYATLVHIVAKMCNLIPGDLIFSGGNTHIYMNHVEQCKEILRREPKELCELVISGLPYKFRYLSTKEQLKYVLKLRPKDFVLNNYVSHPPIKGKMAV
Mouse  NIASYALLTYMIAHITGLQPGDFVHTLGDAHIYLNHIEPLKIQLQREPRPFPKLKILRKV----ETIDDF--------KVEDFQIEGYNPHPTIKMEMAV
Rnorv. NIASYALLTYMIAHITGLQPGDFVHTLGDAHIYLNHIEPLKIQLQREPRPFPKLRILRKV----ETIDDF--------KVEDFQIEGYNPHPTIKMEMAV
Human  NIASYALLTYMIAHITGLKPGDFIHTLGDAHIYLNHIEPLKIQLQREPRPFPKLRILRKV----EKIDDF--------KAEDFQIEGYNPHPTIKMEMAV
Athal. NIASYSLLTCILAHVCDLVPGDFIHVIGDAHVYKNHVRPLQEQLENPPKPFPVLKINPEK----KDIDSF--------VADDFELIGYDPHKKIDMKMAV
Scere. NIASYALLTRMIAKVVDMEPGEFIHTLGDAHVYKDHIDALKEQITRNPRPFPKLKIKRDV----KDIDDF--------KLTDFEIEDYNPHPRIQMKMSV
VZvir. NIAGYALLTYIVAHVTGLKTGDLIHTMGDAHIYLNHIDALKVQLRASPKPFPCLKIIRNV----TDINDF--------KWDDFQLDGYNPHPPLKMEMAL
Pcari. NIASYALLTCMIAHVCDLDPGDFIHVMGDCHIYKDHIEALQQQLTRSPRPFPTLSLNRSI----TDIEDF--------TLDDFNIQNYHPYETIKMKMSI
HVatel NIASYSLLTCMIAHVTDLVPGEFIHTLGDAHVYVNHVDALTEQLTRTPRPFPTLKFMEAV----ASIDDF--------KANDIILENYNPYPSIKMEMAV
Lmajor NIASYALLTILIAKATGLRPGELVHTLGDAHVYRNHVDALKAQLERVPHAFPTLIFKEER----QYLEDY--------ELTDMEVIDYVPHPAIKMEMAV
Calb.  NIASYALLTKMIAHVVDMDCGEFIHTLGDAHVYDKEQFERIPKQFPKLVIKEERKNEIKSIDDF--------KFEDFEIVGYEPYPPIKMKMSV
Pchab. NIASYSIFTYMIAQVCNLQPAEFIHVLGNAHVYNNHVESLKVQLNRTPYPFPTLKLNPEI----KNIEDF--------TISDFTVQNYVHHDKISMDMAA
       *  *      *          *  * * *       *  SITE F    SITE G   *       *
```

B

Fig. 2. **Position of permissive insertion sites along the *E. coli* TS protein.** *A*, multiple alignment of TS sequences from 17 species. Sequences were extracted from the Swissprot Data Bank, and the alignment was performed using the clustal software. *Asterisks* indicate conserved amino acids. *Ecoli, E. coli*; *Bsubt, Bacillus subtilis*; *TN4003*, transposon 4003 (*Staphylococcus aureus*); *Lcasei, Lactobacillus casei*; *Bphi3t*, bacteriophage phi-3T; *BT4*, bacteriophage T4; *Rnorv, Rattus norvegicus*; *Athal, Arabidopsis thaliana*; *Scere, Saccharomyces cerevisiae*; *VZvir*, varicella-zoster virus; *Pcari, Pneumocystis carinii*; *HVatel*, herpesvirus ateles; *Lmajor, Leishmania major*; *Calb, Candida albicans*; *Pchab, Plasmodium chabaudi*. The seven sites (A–G) where insertions have spontaneously occurred in some species were chosen for insertion of HIV protease target sequences in *E. coli* TS. *B*, insertions map. *IS* means insertion. *Numbers* indicate the amino acid immediately preceding the insertions. *E. coli* TS is 264 amino acids long.

Protein concentration was determined in the extracts by the Biuret method (BCA kit, Pierce).

RESULTS

Mapping of Sites Permissive for Insertion of Exogenous Sequences within Thymidylate Synthase—Insertions of a S1 peptide coding sequence (corresponding to a 10-amino acid-long target sequence in the Gag-Pol precursor; see "Materials and Methods") were made at unique restriction sites within the *thyA* gene, but all the mutants thus generated by this method had a thymidylate synthase-deficient phenotype (data not shown). Wild type mutants were also screened in libraries made by insertion of a S1 peptide coding sequence into double stranded breaks created at random position in *thyA* DNA by DNase I digestion. One wild type mutant was obtained by using this procedure (see below). Finally, we analyzed alignments of TS sequences from different species. Although this protein is highly conserved, seven sites can be identified where insertions have spontaneously occurred during the evolution of some species (see Fig. 2 and Ref. 15). We thought that these sites might also accommodate artificial insertions. Thus, seven *E. coli thyA* mutants containing one *Hpa*I restriction site at each of these seven sites were constructed by site-directed mutagenesis. These sites, termed A to G, correspond to different regions of TS from the N to the C terminus, respectively (Fig. 2). Subse-

These phenotypes were analysed in the β-1083 *E. coli* strain grown on minimal medium. ISA–G, insertion sites (Fig. 2). S1 and S4, 10-amino acid-long inserts. S2, 20-amino acid-long insert. S3, 50-amino acid-long insert (see "Materials and Methods" for a description). wt, wild type phenotype. thyA−, thymidylate synthase-deficient phenotype.

	pTZ	pTZ	pTZ	pTZ	pBR
Sequence	S1	S2	S3	S4	S1
ISA,B,C,E,F	mixed	mixed	thyA−	mixed	thyA−
ISD	wt	wt	mixed	wt	mixed
ISG	thyA−	thyA−	thyA−	thyA−	thyA−

quently, these *Hpa*I restriction sites were used to introduce various inserts.

Two double stranded adapters corresponding to two distinct 10-amino acid HIV-1 protease target sequences, termed S1 and S4, were separately introduced into each of these mutants (see "Materials and Methods"). Fourteen mutants were therefore constructed. The phenotypes of these mutants, carried by the high copy number pTZ 18 plasmid, were determined in the β-1083 *E. coli* strain grown on minimal medium (see "Materials and Methods"). Three phenotypes were obtained (Table I). Insertions at site G caused a thymidylate synthase-deficient phenotype. This site is probably too close to the C terminus, which has been shown to be critical for thymidylate synthase activity (Ref. 21 and references therein). Mutants carrying an insert at sites A, B, C, E, and F displayed a "mixed" phenotype characterized by the capacity to grow on both negative and positive selection media. These mutants are thus able to synthesize enough dTMP to ensure bacterial growth in the absence of exogenous thymidine. They have, however, a thymidylate synthase activity sufficiently reduced to avoid growth inhibition in the presence of trimethoprim. Finally, only mutants carrying an insert at site D displayed an unambiguous wild type phenotype. The wild type mutant isolated after random insertion of S1 into *thyA* is also located in this region. The phenotypes of these three types of mutants seem to be determined mainly by the position and not the sequence of the insertion, because similar phenotypes were obtained with the two inserted sequences.

Extent of Thymidylate Synthase Plasticity—We then evaluated the maximum length of peptides that could be inserted into TS. In this heterologous context, increasing the insert size might increase the efficacy of cleavage by the HIV-1 protease or decrease the activity of uncleaved TS, thereby facilitating the switch to a thymidylate synthase-deficient phenotype. For this purpose, an oligonucleotide termed S2, corresponding to a 20-amino acid substrate peptide encompassing the S1 target sequence (see "Materials and Methods"), was introduced into *thyA* at each of the seven *Hpa*I sites. The phenotypes of these new mutants are similar to those of the mutants carrying the 10-amino acid coding insert (Table I), indicating that doubling the size of the insert had a little effect on the activity of the proteins. A 50-amino acid coding sequence, termed S3 (see "Materials and Methods"), was next inserted into *thyA*. Except insertions at site D, which induced a mixed phenotype, all the other mutants display a thymidylate synthase-deficient phenotype (Table I). Thus the permissiveness of TS seems to be limited to 50-amino acid-long insertions.

In order to construct an *E. coli* strain with a thymidylate synthase-deficient phenotype conferred by HIV-1 protease activity, it is important to know the extent of TS destruction that must be achieved. pTZ 18 is a high copy number plasmid (approximately 500 copies/cell). All of the mutants bearing the 10-amino acid-long S1 insert were subcloned into the low copy number pBR plasmid (approximately 20 copies/cell), leading in

theory to at least a 25-fold reduction in the production of TS protein. As shown in Table I, all of the mutants that had a mixed phenotype when carried by the pTZ vector had a thymidylate synthase-deficient phenotype when carried by the pBR vector. Those bearing an insert at site D, which had a wild type phenotype when carried by pTZ 18, had a mixed phenotype when carried by pBR. These results show that the TS mutants have a reduced thymidylate synthase activity. They also demonstrate that it is possible to change the phenotype of *E. coli* by adjusting, within reasonable limits, the level of expression of the mutants. In fact, subcloning the mutant sequences from pTZ 18 into pBR 322 reduced the level of thymidylate synthase activity below the threshold necessary for growth on a positive selection medium, with the exception of mutants with an insert at D site.

Taken together, the results of the insertional mutagenesis of *thyA* show that site D is more permissive for the insertion of exogenous sequences than any other site we have identified.

Finally, to test the plasticity of TS further, we made two insertions in the same molecule, each of them at a different permissive site. This could also improve TS cleavage by the HIV-1 protease and/or modulate the activity of TS. For this purpose, we constructed seven "double mutants" that carry two HIV-1 target sequences. All these double mutants have a thymidylate synthase-deficient phenotype (data not shown). The structure of TS proteins with two insertions is probably too profoundly perturbed to maintain the integrity of the catalytic site.

Effect of Insertions on the Activity of Thymidylate Synthase—In order to get a more quantitative estimate, thymidylate synthase activity was measured in crude extracts of proteins prepared from bacteria expressing the different *thyA* mutants. (see "Materials and Methods"). Twelve *thyA* mutants bearing S1 or S4 inserts were tested. They all had a very low activity estimated to be <1% compared with the wild type. This loss of activity must be attributed to a reduction in enzymatic activity of the mutated enzymes rather than to a reduction in their half-life. Indeed, immunoblotting experiments (see below) did not show a reduction in the quantity of TS expressed by the *thyA* mutants. These *thyA* mutants were overexpressed because of the p*lac* promoter and because pTZ 18 vector is a high copy number plasmid (500 copies/cell). Thus, it is not surprising that in spite of their highly reduced activity, they ensured growth on the positive selection medium. Moreover, other mutants with <1% activity relative to wild type have already been shown to grow on the positive selection medium (13, 22).

In Vivo Cleavage of Insertions by HIV-1 Protease—Protein extracts were prepared from *E. coli* cells expressing the various *thyA* mutants with or without HIV-1 protease (*prt*+ or *prt*− respectively; see "Materials and Methods") and analyzed by immunoblotting with TS-specific antibodies (Fig. 3). In all the tested mutants, the 30.5-kDa TS band showed a high reduction in intensity in cells expressing HIV-1 protease, whereas this is not the case with wild type TS used as a control (Fig. 3*C*). In some experiments the cleavage products were not detected (Fig. 3*A*), probably due to their rapid degradation, as is often the case in *E. coli* (11, 23). However, bands of 21 and 20 kDa, corresponding to one of the two expected cleavage products, could be seen in several other experiments with *ISB* and *ISC* mutants, respectively (Fig. 3, *B* and *C*). The other cleavage product was not detected; either the product was not recognized by the antibodies or it was degraded. These data indicate that HIV-1 protease is able to recognize and cleave, in this heterologous context, the two target sequences at various positions within the TS protein.

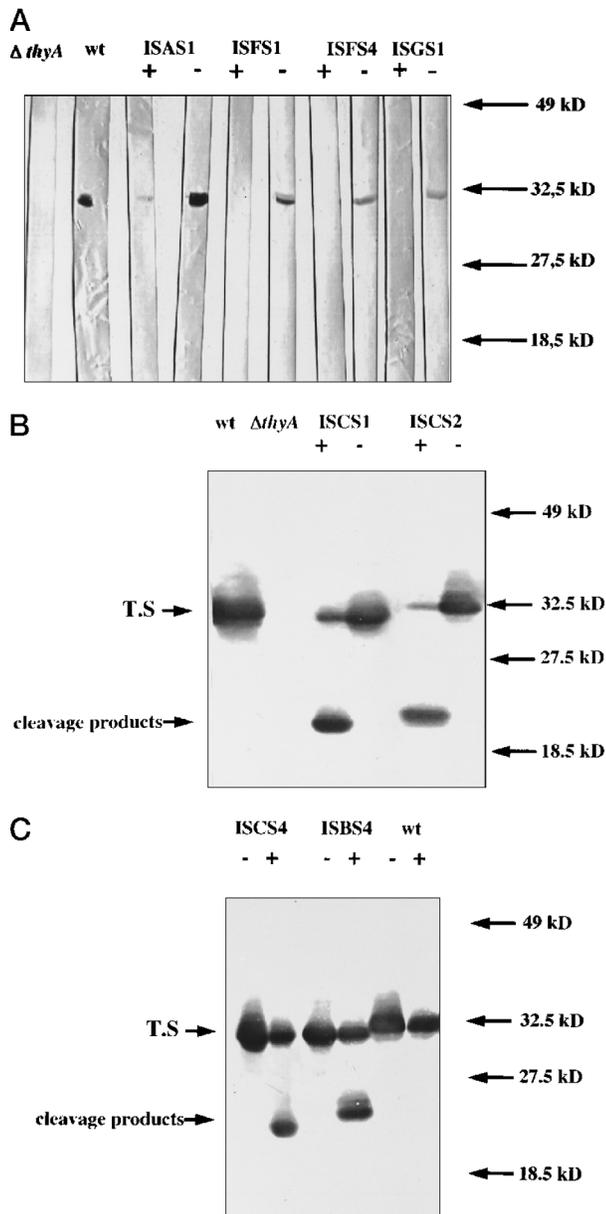*Thymidylate Synthase Cleavage by HIV-1 Protease Confers a*

FIG. 3. *In vivo* **cleavage of insertions within thymidylate synthase.** β-1083 cells were transformed by two constructs, a pTZ *thyA* mutant and a pSU *prt* plasmid expressing either an active (+) or a defective (−) HIV protease protein. For each *thyA* mutant, as well as for *thyA* wild type, the quantity of cells was normalized in *prt*+ and *prt*− strains. Proteins were extracted (see "Materials and Methods") and analyzed by immunoblotting using antibody A (*A*) or B (*B* and *C*).

Thymidylate Synthase-deficient Phenotype—Finally, we analyzed the phenotypes of Δ-*thyA E. coli* strains cotransfected with a plasmid expressing one of the *thyA* mutants and a second plasmid expressing either active HIV-1 protease or, as a control, the same construct with a deleted protease. A first series of experiments carried out on minimal medium with the β-1083 strain gave negative results. Because β-1083 is a very slow growing strain, we repeated the experiments in a faster strain for which the selection based on thymidylate synthase activity should be more efficient. β-1308 cells were transformed with the same plasmids as in the first series of experiments, and phenotypes were analyzed on Müeller-Hinton medium. Under these conditions β-1308 cells grew four times faster than β-1083 on minimal medium. In this context *ISES1* and *ISFS1* mutants exhibited a protease-dependent phenotype on the negative selection medium. On the plus medium the plating effi-

TABLE II
HIV-1 protease-dependent phenotypes of thyA mutants in the β-1308 E. coli strain

Cells were cotransformed by two plasmids, one carrying either an active or a defective HIV-1 protease gene cloned in a pTZ vector and another expressing a *thyA* mutant cloned in a pSU vector. These cells were plated on positive and negative selection media (called plus and minus in this table, respectively) as well as nonselective medium and titrated by serial dilutions. The values are the numbers of colonies growing on the different selective media, relative to the number of colonies growing on the nonselective medium. thyA−, thyA deleted. thyA+, wild type. ND, not determined.

	plus medium	minus medium
	%	
prt+, ISES1	ND	42
prt−, ISES1	ND	0.3
prt+, ISFS1	ND	70
prt−, ISFS1	ND	0.3
prt+, thyA−	<0.001	100
prt−, thyA−	<0.001	100
prt+, thyA+	100	<0.001
prt−, thyA+	100	<0.001

ciency was very low and prevented us from estimating growth differences. This could be due to the combination of two factors: the reduced viability of cells with a *thyA* mutant and the toxicity of HIV protease in *E. coli* (11, 24). As illustrated in Table II, there was a 230- and 140-fold induction of *prt*+ *ISFS1* and *prt*+ *ISES1* cell growth on the negative selection medium. The other mutants did not exhibit the thymidylate synthase-deficient phenotype, probably because the levels of thymidylate synthase activity and/or cleavage by the protease were insufficient.

DISCUSSION

Comparison of TS sequences from different species enabled us to identify six sites permissive for insertions within the TS protein. We were able to insert 20 or even 50 amino acids in the case of site D, showing that insertions at these sites allow a residual enzymatic activity that is still sufficient for bacterial growth. These results are in agreement with crystallographic data that show that the natural insertions that occurred spontaneously during the evolution of some species correspond to surface loops and do not contribute to the active site (15, 16). All the insertions made in other regions of TS protein failed to produce an active TS. Because TS is a highly structured and conserved protein, it is likely that the insertions markedly disturbed the core structure of the protein and therefore abolished enzymatic activity. We also observed the *in vivo* cleavage of TS by HIV-1 protease for two different target sequences inserted at the six different permissive sites, whereas wild type TS was unaffected. This result demonstrates that these insertions are accessible to HIV protease and that cleavage of the natural substrate sequences can occur in this heterologous context. *In vivo* cleavage in a heterologous context has already been reported for the S1 sequence (9) but not S4. It is remarkable that cleavage occurred in six different contexts, suggesting that accessibility is the only parameter important for cleavage besides the target sequence itself. This cleavage conferred a protease-dependent phenotype for two of these mutants. In the other mutants the extent of TS cleavage is probably too low for phenotypic switch. A nonspecific toxic effect of HIV-1 protease on *E. coli* growth has been reported (11, 24). This nonspecific toxic effect of HIV-1 protease cannot account for growth induction on the negative selection medium. *ISFS1* and *ISES1 prt*− cells exhibited background growth between 4×10^{-3} and 7×10^{-3}, respectively, on the negative selection medium. These values are in the same range as the background growth reported in another study using HIV-1 protease and the tetracy-

cline resistance protein as a target (10) but higher than the 10^{-6} background observed with the zucchini yellow mosaic virus protease (12). The background in our TS system could probably be considerably reduced by optimizing the ratio of expression of TS relative to HIV-1 protease and therefore improving the rate of TS cleavage.

At a more general level our data show that it should be possible to induce a selectable phenotype in *E. coli* through the protease-mediated destruction of an essential metabolic enzyme and therefore that the genetic selection of bioactive molecules from large libraries of peptides is feasible. *E. coli* strains with a protease-dependent growth phenotype have previously been described (10–12). In these systems the proteases induce antibiotic-sensitive phenotypes that provide a means of selecting protease inhibitors, as is the case with TS on the positive selection medium (Fig. 1). In this type of selection the destruction of a target by a protease inhibits bacterial growth. As we have shown in this study, TS destruction induces bacterial growth on the negative selection medium. Because of this property, our engineered *E. coli* strain could be used not only for the selection of protease inhibitors but also for the study of protease specificity.

REFERENCES

1. Beynon, R. J., and Bond, J. S. (eds) (1994) *Proteolytic Enzymes,* IRL Press at Oxford University Press, Oxford
2. Scharpe, S., De Meester, I., Hendriks, D., Vanhoof, G., van Sande, M., and Vriend, G. (1991) *Biochimie (Paris)* **73,** 121–126
3. Hellen, C. U. T., and Winner, E. (1992) *Curr. Opin. Biotechnol.* **3,** 643–649
4. Tewari, M., Quan, L. T., O'Rourke, K., Desnoyers, S., Zeng, Z., Beidler, D. R., Poirier, G. G., Salvesen, G. S., and Dixit, V. M. (1995) *Cell* **81,** 801–809
5. Fitzgerald, P. M. D., and Springer, J. P. (1991) *Annu. Rev. Biophys. Biophys. Chem.* **20,** 299–320
6. Tomasselli, A. G., and Heinrikson, R. L. (1994) *Methods Enzymol.* **241,** 279–301
7. Farmerie, W. G., Loeb, D. D., Casavant, N. C., Hutchison, C. A., Edgell, M. H., and Swanstrom, R. (1987) *Science* **236,** 305–308
8. Debouck, C., Gorniak, J. G., Strickler, J. E., Meek, T. D., Metcalf, B. W., and Rosenberg, M. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 8903–8906
9. Baum, E. Z., Bebernitz, G. A., and Gluzman, Y. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 10023–10027
10. Block, T. M., and Grafstrom, R. H. (1990) *Antimicrob. Agents Chemother.* **34,** 2337–2341
11. McCall, J. O., Kadam, S., and Katz, L. (1994) *Bio/Technology* **12,** 1012–1015
12. Balint, R. F., and Plooy, I. (1995) *Bio/Technology* **13,** 507–510
13. Belfort, M., and Pedersen-Lane, J. (1984) *J. Bacteriol.* **160,** 371–378
14. Stacey, K. A., and Simson, E. (1965) *J. Bacteriol.* **90,** 554–555
15. Hardy, L. W., Finer-Moore, J. S., Montfort, W. R., Jones, M. O. Santi, D. V., and Stroud, R. M. (1987) *Science* **235,** 448–454
16. Matthews, D. A., Villafranca, J. E., Janson, C. A., Smith, W. W., Welsh, K., and Freer, S. (1990) *J. Mol. Biol.* **214,** 937–948
17. Belfort, M., Maley, G., Pedersen-Lane, J., and Maley, F. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 4914–4918
18. Martinez, E., Bartolome, B., and dela Cruz, F. (1988) *Gene (Amst.)* **68,** 159–162
19. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) *Methods Enzymol.* **154,** 367–382
20. Roberts, D. W. (1966) *Biochemistry* **5,** 3546–3548
21. Carreras, C. W., Costi, P. M., and Santi, D. V. (1994) *J. Biol. Chem.* **269,** 12444–12446
22. Michaels, M. L., Kim, C. W. Matthews, D. A., and Miller, J. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 3957–3961
23. Goldberg, A. L. (1972) *Proc. Natl. Acad. Sci. U. S. A.* **69,** 422–426
24. Baum, E. Z., Bebernitz, G. A., and Gluzman, Y. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 5573–5577