FAST TRACK

Identification of the minimal conserved structure of HIV-1 protease in the presence and absence of drug pressure

Francesca Ceccherini-Silberstein[a], Fulvio Erba[a], Federico Gago[b], Ada Bertoli[a], Federica Forbici[c], Maria Concetta Bellocchi[c], Caterina Gori[c], Roberta d'Arrigo[c], Luisa Marcon[d], Claudia Balotta[e], Andrea Antinori[c], Antonella d'Arminio Monforte[e] and Carlo-Federico Perno[c]

Objective: To define the extent of amino acid protease (PR) conservation *in vivo* in the absence and presence of pharmacological pressure in a large patient cohort.

Methods: Plasma-derived complete protein PR sequences from a well-defined cohort of 1096 HIV-1 infected individuals (457 drug-naive and 639 under antiretroviral therapy including PR-inhibitors) were obtained and analysed, and are discussed in a structural context.

Results: In naive patients, the PR sequence showed conservation (< 1% variability) in 68 out of 99 (69%) residues. Five large conserved regions were observed, one (P1–P9) at the N-terminal site, another (E21–V32) comprised the catalytic active-site, a third (P44–V56) contained the flap, a fourth contained the region G78–N88, and another (G94–F99) contained the C-terminal site. In PR-inhibitor treated patients, the appearance of mutations primarily associated with drug resistance determined a decrease of amino acid invariance to 45 out of 99 residues (45% conservation). The overall degree of enzyme conservation, when compared to the PR sequences in drug-naive patients, was preserved at the N- and C-terminal regions, whereas the other large conserved areas decreased to smaller domains containing, respectively, the active-site residues D25–D29, the tip of the flap G49–G52, and the G78–P81 and G86–R87 turns.

Conclusions: Amino acid conservation in HIV PR can be minimally present in 45 residues out of 99. Identification of these invariable residues, with crucial roles in dimer stability, protein flexibility and catalytic activity, and their mapping on the three-dimensional structure of the enzyme will help guide the design of novel resistance-evading drugs.
© 2004 Lippincott Williams & Wilkins

AIDS 2004, **18**:11–19

Keywords: HIV drug resistance, protease, genotype, conservation, antiretroviral therapy, resistance mutations

From the [a]Department of Experimental Medicine, University of Rome 'Tor Vergata', Italy, the [b]Department of Pharmacology, School of Medicine, University of Alcala, Spain, [c]National Institute for Infectious Diseases, 'L. Spallanzani', Rome, the [d]Department of Histology, Microbiology and Medical Biotechnologies, University of Padua, and the [e]Institute of Infectious and Tropical Diseases, University of Milan, Italy.

Correspondence to C.F. Perno, National Institute for Infectious Diseases, 'Lazzaro Spallazani' IRCCS, Via Portuense 292, 00149 Rome, Italy.

Received: 30 November 2004; revised: 29 April 2004; accepted: 7 May 2004.

DOI: 10.1097/01.aids.0000131394.76221.02

ISSN 0269-9370 © 2004 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

Introduction

The seemingly unlimited generation of new viral variants is a hallmark of HIV-1 spread among humans [1–3]. This intrinsic variability is further enhanced by the selective pressure imposed by pharmacological treatments, resulting in mutations particularly in the *pol* region encoding reverse transcriptase (RT) and protease (PR) [4–7]. HIV-1 PR is an aspartic protease consisting of two identical 99-amino acid monomers, whose association is essential for processing the viral *gag–pol* polyprotein precursors into mature structural and enzymatic proteins and viral infectivity. As a consequence, HIV-1 protease inhibitors (PI) are an important class of potent antiretroviral agents [8–11].

Numerous studies have contributed to our current knowledge of HIV-1 PR polymorphisms and drug-related variants [6,12–26]. To date, mutations at 50 residues in the PR sequence have been related to treatment with one or more experimentally tested PI [25,27–29]; of these, 22 are involved in resistance against the six currently approved PI: saquinavir, ritonavir, indinavir, nelfinavir, amprenavir and lopinavir [27,30]. Most of these studies have focused on the prevalence of mutations associated with PI resistance.

The alternative perspective of focusing on those amino acids that are highly conserved has received comparatively little attention and it has been limited to a few studies, mostly based upon a small number of sequences [31–33]. Only recently, description of protein conservation and mutational patterns of HIV-1 PR was obtained from sequences derived from a large group of both drug-naive and drug-treated patients [25].

Due to the extensive polymorphism and drug-related variability in HIV-1 PR, a large amount of data will be required to optimally define the minimal conserved structure of this enzyme. To facilitate this effort, we now extend previous observations by comparing viral PR sequences obtained from a well-defined cohort of 1096 HIV-1 infected individuals. In a further attempt to define an improved minimal conserved structure of HIV-PR, we also rationalize our findings in a structural elaboration, as this understanding is important for the design of new resistance-evading PI.

Materials and methods

Patients

The study included 457 patients naive for antiretroviral drugs and 639 patients failing highly active antiretroviral therapy (HAART) regimens containing at least one PI, and who where enrolled either in the Italian Cohort of Antiretroviral Naive patients (I.CO.N.A.) or in different clinical centres in central Italy. Overall, drug-treated patients were exposed to a median of two PI (279 patients to one PI, 201 to two PI, 97 to three PI, and 62 to four or more PI). Indinavir was used in 63.2% of patients, nelfinavir in 51.8%, saquinavir in 41.9%, ritonavir in 29.4% (with 8% at a low dose as part of a dual-PI combination), lopinavir in 4.2%, and amprenavir in 1.1%. At the time of genotypic analysis, 36.9% of patients were under treatment with nelfinavir (with a median time of 463 days), 23.3% with indinavir (median time, 718 days), 10.2% with ritonavir (median time, 278 days), 9.5% with saquinavir (median time, 611 days), 3.8% with lopinavir/ritonavir (median time, 297 days), and 0.8% with amprenavir (median time, 316 days).

HIV-subtype analysis showed that 18 out of 457 (3.9%) drug-naive patients carried non-clade B pol subtypes (F, four; A, two; C, one; G, one; recombinant subtypes, ten). Similarly, 20 out of 639 PI-treated patients (3.1%) carried non-clade B pol subtypes (F, six; C, two; G, one; D, one; recombinant subtypes, ten).

HIV sequencing

HIV genotype analysis was performed on plasma samples by means of a commercially available HIV genotyping kit (the ViroSeq HIV-1 Genotyping System, AB [Celera Diagnostics/Abbott Molecular Diagnostics, Roma, Italy]). In brief, RNA was extracted, retrotranscribed by murine leukaemia virus RT, and amplified with amplitaq-Gold polymerase enzyme by using two different sequence-specific primers for 40 cycles [20,21]. *Pol* amplified products (containing the entire PR and the first 320 amino acids of the RT open reading frame) were sequenced full-length in both the sense and antisense orientations by using seven different overlapping sequence-specific primers and an automated sequencer (ABI 3100 [Applied Biosystems, Foster City, CA, USA]) [20,21]. For each individual, the quality endpoint was ensured by a coverage of the entire PR sequence by at least three sequence segments. Sequences having a mixture of wild-type and mutant residues at single positions were considered to have a mutation at that position. When the mixture was between two different mutations, both mutations were considered and reported (see Fig. 2).

The majority of nucleotide sequences of drug-naive patients have already been submitted to GenBank [21]; the others are in the process of being submitted.

Mutations

Consensus B was used as a reference strain for the definition of mutations. All mutations associated (by *in vitro* or *in vivo* studies) with resistance to PI currently discovered [27,30], as well as mutations not yet associated with drug resistance, were analysed. All 99 amino acids of PR were screened in both drug-naive

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

and PI-treated patients, and the frequency of mutations was calculated and statistically compared using the chi-square test (based on a 2×2 contingency table containing the numbers of isolates from untreated and treated persons, and the number of isolates with and without mutations).

Amino acids/regions not mutated or mutated with $\leqslant 1\%$ prevalence over each cohort of patients (drug-naive or treated) were defined as conserved. Conservative and non-conservative amino acid substitutions were recognized according to ClustalW [34].

Structural analysis

To facilitate visualization of HIV-PR conservation, in both naive and treated patients, all amino acids were mapped onto a three-dimensional representation of the enzyme and colour-coded according to their mutation frequency rate using the Swiss PDB viewer v.3.7 software [35] and the x-ray coordinates deposited in the Protein Data Bank (PDB; http://www.rcsb.org/PDB/) with code 1HIH [36].

Results

Degree of conservation of HIV-1 PR

HIV-1 PR conservation *in vivo* both in the absence and presence of PI pressure was assessed by evaluating the entire protein sequences derived from 457 drug-naive and 639 HAART–PI-treated patients.

The analysis of sequences from drug-naive patients showed conservation ($\leqslant 1\%$ variability) in 68 out of 99 amino acids (69% overall conservation) (Fig. 1). Some

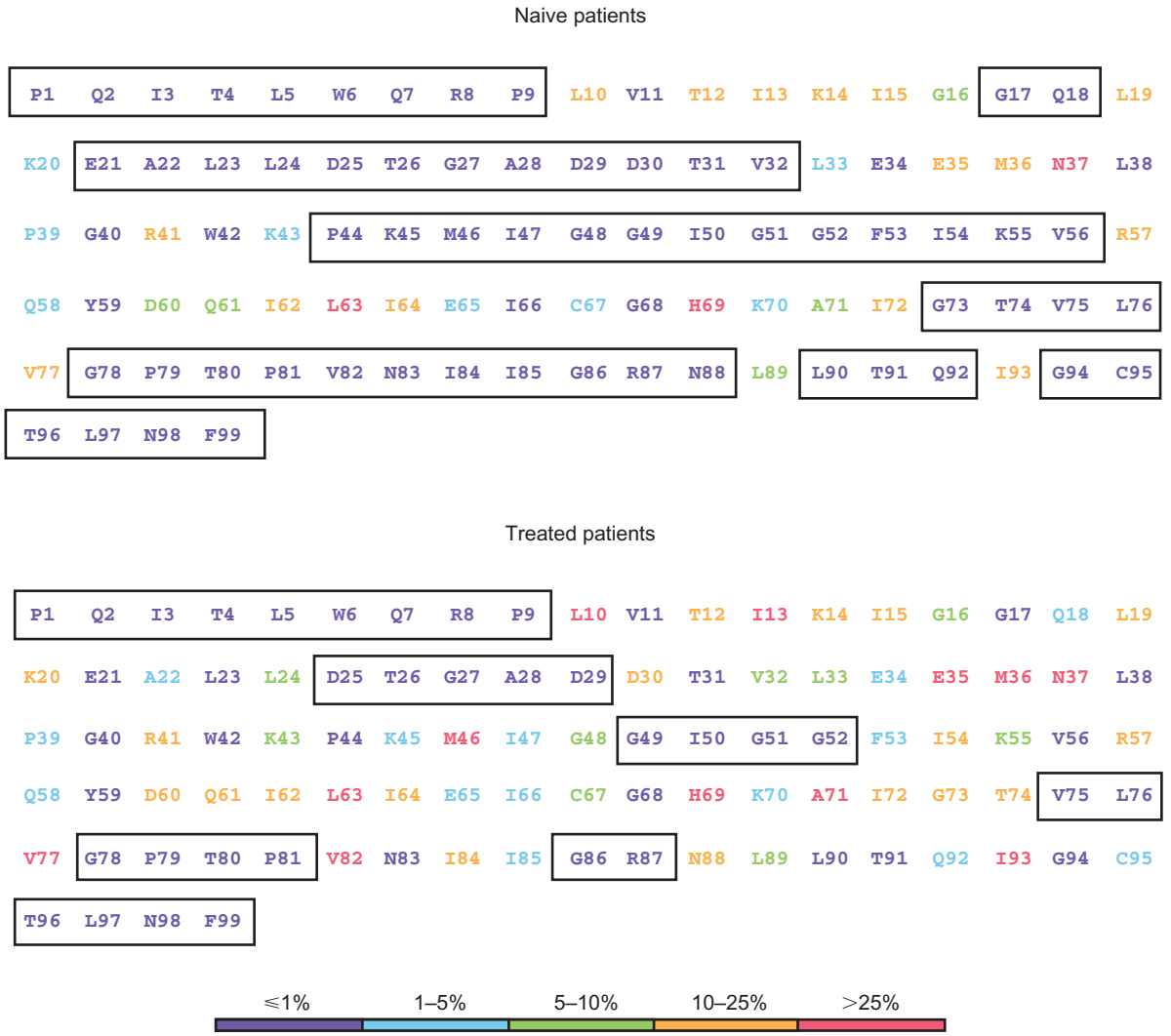

Fig. 1. Conserved regions of HIV-1 PR in drug-naive and drug-treated HIV-1 infected patients. The amino acid sequence of HIV-1 PR (99 amino acids) of clade B consensus (shown as a reference) is coloured according to the frequency rate of mutations observed in plasma samples from 457 drug-naive and 639 PI-treated patients. Conserved domains or stretches of amino acids are boxed. The bar indicates the frequency rate of mutations (%) relative to the colours used in the figure.

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

invariant residues were scattered (V11, E34, L38, G40, W42, Y59, I66, G68), while others clustered to form discrete conserved domains, that varied in size between the untreated and treated groups (Fig. 1, boxed areas). Five large non-contiguous conserved PR regions were defined in the naive patients' group. For the most part, these contain functionally important domains, such as the N and C termini involved in dimerization (P1–P9; G94–F99), the active-site region (E21–V32), the flap (P44–V56), and the G78–N88 region (Fig. 1). In addition, three small conserved stretches of two to four amino acids each (G17–Q18, G73–L76, L90–Q92) were observed.

The analysis of PR protein sequences from drug-treated patients showed a different picture, characterized by a substantial decrease of amino acid conservation, mainly (but not exclusively) due to the emergence of major and minor PI resistance mutations (Fig. 1). Conserved amino acids were observed at 45 out of 99 residues (45% of conservation versus 69% in naive patients), and several of them clustered in discrete areas. Compared to drug-naive patients, the degree of conservation in drug-treated patients was identical at the N- (P1–P9) and C-terminal (G94–F99) regions, with the exception of residue C95, for which the variation frequency increased from 0.6% to 2.7%. All the other large conserved areas (E21–V32, P44–V56, G78–N88) shrank to smaller domains containing, respectively, the active-site residues D25–D29, the tip of the flap G49–G52, and the G78–P81 and G86–R87 turns. The three small conserved stretches (G17–Q18, G73–L76, and L90–Q92) disappeared, leaving only a few and scattered highly-conserved amino acids (maximally two in one case).

Pattern of HIV-1 PR mutations

Several PR polymorphisms were present in drug-naive patients. Among 31 variable residues, 23 were mutated in $> 5\%$ of drug-naive patients and three of them (N37, L63, and H69) were highly variable (substituted in $> 25\%$) (Fig. 1). Polymorphic changes included several known compensatory PI-resistance mutations, such as L10I, M36I, N37D, L63P, A71T, V77I, L89M, I93L (Fig. 2). By contrast, the frequency of all major drug-resistance mutations was consistently $\leqslant 1\%$.

The development of drug resistance during therapy failure resulted in a general increase of mutations in the PR from PI-treated patients, with 40 out of 99 residues found mutated in $> 5\%$ of patients (compared to 23 in drug-naive patients), of which 14 were highly variable (substituted in $> 25\%$) (Fig. 1). Statistically significant differences (with χ^2 test; $P < 0.05$) in the frequency of variability between drug-naive and PI-treated patients were found at 38 out of 99 residues (shown in boldface type in Fig. 2). At these 38 positions we observed some classical PI resistance associated mutations, two substitutions not yet linked to PI treatment (T12 and Q61) and 14 substitutions that have been only recently linked to PI treatment (A22, E34, E35, K43, K45, K55, Q58, I62, I66, I72, T74, I85, Q92, C95) [25]. The frequency of variability at these 38 residues was increased in the group of PI-treated patients, with a single exception at position T12, where mutability decreased from 18.2% in the drug-naive group to 13.6% in the PI-treated group ($P = 0.049$).

Some unusual changes at resistance-associated sites (K20I, K20T), as well as changes at residues not yet associated with resistance to any known PI (E35D, K43T, K45R, K55R, Q58E, Q61H, I62V, I72T, T74S, I85V, Q92K, C95F), had a significantly increased representation ($P < 0.05$ to < 0.001) in treated patients. In contrast, the rate of amino acid substitutions at few sites was significantly ($P < 0.01$) diminished in treated patients as compared to drug-naive patients. Among these, the T12A mutation decreased from 4.8 to 1.4%, the L63Q mutation from 4.2 to 1.1%, L63S from 6.1 to 3.3% and H69N from 12 to 7.4%. However the prevalence of wild-type alleles at positions 63 and 69 was not increased in treated patients, while at position 12 the wild-type allele showed increased frequency (Fig. 2).

It is interesting to note that all of the most frequent amino acid mutations, both in naive and in treated patients, are conservative in terms of charge/polarity. Few exceptions are at residues 63 (L63P, $> 25\%$; L63S/T/Q, 3–6%), 16 (G16E, $> 4\%$), 72 (I72T, $> 3\%$), and 20 (K20I, 2.4% in naive patients; K20I/M/T, 3–7% in PI-treated patients). The following non-conservative mutations are found only in the treated patients' group: G73S ($> 13\%$), G48V (4.7%) and C95F (2.3%).

Structural model of the minimal conserved areas of HIV-1 PR

The PR from naive patients retains the overall structure of the active-site that is seen in the wild-type enzyme.

Fig. 2. Polymorphisms and mutations in HIV-1 PR from drug-naive and PI-treated patients. Amino acid sequence mutations in plasma samples of HIV-1 PR from 457 drug-naive and 639 drug-treated patients are reported. The clade B consensus sequence is shown as a reference. Residues associated with PI-treatment by previous *in vitro* or *in vivo* studies [25,27,29] are underlined. Residues whose wild-type prevalence differs significantly ($P < 0.05$) between naive and treated patients are shown in boldface type. The predominant mutation for each position is given closest to the reference. Mutations are represented by one-letter amino acid symbols and are numbered.

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

Naive	Consensus	Treated
	$N_1D_1A_1$P1	
	R_1Q1K_2P_1	
	I3V_2	
	S_2T4S_5P_3	
	R_1L5	
	R_1W6C_1G_1	
	H_1Q7E_1R_1	
	R8Q_1D_1	
	L_1P9L_1	
	$F_1V_{20}I_{33}$**L10**$I_{250}V_{46}F_{24}M_5$	
	I_2V11I_3A_1	
	$K_1D_1V_1E_2I_4N_6P_{20}A_{23}S_{26}$**T12**$P_{25}S_{23}N_{10}A_{10}I_5E_3V_2D_1K_1$	
	A_3V_{100}I13V_{177}	
	E_3R_{74}K14$R_{101}M_3E_2I_2Q_2H_1$	
	L_1V_{101}I15$V_{123}L_1$	
	A_3E_{30}G16$E_{28}A_{10}$	
	E_6G17D_2	
	$N_1E_1L_2$Q18$H_{10}E_4V_1Y_1$	
	$A_1K_1H_1S_1E_2P_2T_5Q_8V_9I_{31}$L19$I_{43}V_{17}Q_{15}T_6P_3A_1E_1F_1$	
	$M_1R_6I_{11}$**K20**$R_{61}I_{43}M_{28}T_{24}V_4L_1$	
	K_1E21Q_1	
	A22$V_{10}T_1$	
	L23$I_3Q_1V_1$	
	F_1**L24**$I_{42}F_4$	
	D25Y_3	
	I_2T26	
	G27	
	V_1A28G_1	
	D29V_1	
	D30$N_{71}Y_1$	
	T31A_1	
	V32$I_{34}E_1$	
	$I_2F_3V_7$**L33**$F_{19}I_{13}V_8P_1$	
	E34$Q_7K_4D_2$	
	K_1D_{96}**E35**$D_{168}G_{12}N_4Q_1$	
	$N_1T_1V_2L_6I_{101}$**M35**$I_{194}V_{17}L_6$	
	$R_1H_4Y_7E_7A_7C_{11}T_{11}D_{54}S_{58}$**N37**$D_{116}S_{45}E_{18}T_{10}K_7C_4H_4Y_3A_1G_1I_1P_1W_1$	
	W_1F_1L38F_1	
	$L_1T_1S_5$P39S_8Q_2	
	G40	
	I_1K_{107}R41$K_{126}N_2$	
	C_1W42	
	T_1R_{13}**K43**$T_{25}R_{14}N_3I_2E_1S_1$	
	P44	
	Q_1R_2**K45**$R_{20}Q_2T_1L_1$	
	L_4**M46**$I_{153}L_{50}V_4$	
	I47$V_{15}L_1A_1$	
	G48$V_{30}M_2I_1Q_4$	
	G49	

Naive	Consensus	Treated
	I50	
	G51	
	G52	
	F53$L_{22}Y_2I_1$	
	I54$V_{109}A_8T_6L_4S_3$	
	R_3**K55**$R_{31}H_3N_2$	
	V56	
	$Q_1S_2K_{49}$**R57**K_{72}	
	H_1E_6**Q58**$E_{29}S_1$	
	Y59	
	E_{40}D60E_{67}	
	$S_1H_1D_5N_7E_{13}$**Q61**$E_{25}H_{19}N_{13}D_9$	
	L_1V_{105}**I62**$V_{234}S_2L_1$	
	$G_1N_1F_2I_2V_4R_5C_7H_{16}A_{16}T_{17}Q_{19}S_{29}P_{241}$**L63**$P_{462}T_{38}A_{26}S_{14}Q_8H_4C_4V_2I_1N_1$	
	$L_8M_8V_{89}$I64$V_{113}L_{16}M_{10}$	
	D_{20}E65$D_{23}Q_3H_1K_1$	
	I66$F_7L_1V_4$	
	D_8Y_9C67$Y_{13}F_8E_7D_6S_3H_1$	
	E_1G68E_3A_1	
	$I_1R_1Q_5Y_Y N_{55}K_{62}$H69$K_{84}N_{51}Q_{16}Y_9R_5T_3I_2S_2$	
	$E_1T_1R_8$K70$R_{13}T_2E_2$	
	V_8T_{26}**A71**$V_{211}T_{79}I_{18}L_1$	
	$M_1L_2E_4T_{15}V_{30}$**I72**$V_{58}T_{45}M_{12}L_{12}R_{10}E_5K_5A_3$	
	G73$S_{86}T_{12}C_{11}A_3K_2V_2$	
	A_1**T74**$S_{41}A_{18}P_{12}E_1$	
	E_1V75$I_4T_1A_1$	
	L76V_6	
	I_{113}**V77**I_{223}	
	G78	
	$L_1D_1S_2$P79$S_2A_2D_2$	
	T80	
	P81	
	I_6**V82**$A_{146}T_{18}I_9F_7C_3S_2$	
	s_1N83D_5	
	I84$V_{72}C_1$	
	V_1**I85**$V_{24}F_1M_1$	
	G86E_1	
	R87	
	N88$D_{55}S_8T_3$	
	M_{29}L89$M_{27}V_7I_7$	
	L90$M_{255}V_1$	
	A_1T91I_2A_1	
	$R_1E_1K_1$**Q92**$K_{16}R_6N_1$	
	M_1L_{86}**I93**$L_{213}M_5V_1$	
	S_1G94	
	$A_1F_1W_1$**C95**$F_{15}R_1L_1$	
	T96A_1P_1	
	L97	
	N98K_1	
	L_1**F99**$S_3L_1V_1$	

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

and are not widely dispersed over the protein structure. Most are located in loops that expose the side chains to the solvent, suggesting that polymorphism may be selected also by cellular factors possibly interacting with viral proteins [11].

The coloured surface model of the PR from drug-treated patients, on the other hand, shows a patchwork pattern representing a mosaic of mutations that reflects the enzyme's ability to evolve in the presence of drug pressure. The overall surface is more 'patchy' compared to that of the naive patients' model, although a conserved region (blue) is present along the median-line, representing a stretch that is essential for dimer formation and stabilization (Fig. 3a). Many variable residues are displayed in peripheral areas that are widely dispersed over the protein structure. In contrast to what is observed in naive patients, the binding cavity displays a high and variable occurrence of mutations at positions 30, 32, 48, 82 and 84 (Fig. 3b), which are all found lining the walls of the substrate/inhibitor-binding site.

Finally, it is noteworthy that a low mutational rate (around 5%) is present in our cohort of treated patients in regions found to be conserved in naive patients. These areas (around the active site and the substrate/inhibitor-binding site) are typically highly associated to resistance to specific PI [27]. The frequency of the L24I mutation (specific for indinavir, lopinavir/ritonavir and atazanavir), L33F mutation (specific for ritonavir, lopinavir/ritonavir, atazanavir and tipranavir), as well as V32I (multi-PI resistance mutation) and G48V (specific for saquinavir) are low. Indeed the frequency of these mutations in our study is much lower than expected, when considering the adequate pharmacological exposure of the patients' cohort (63.2% indinavir, 41.9% saquinavir, 29.4% lopinavir, 4.2% lopinavir/ritonavir).

Discussion

The present study presents an improved definition of the minimal unchanged areas of HIV-1 PR that might be essential for the preservation of a stable and functional enzyme and therefore represent target sites in the design of novel resistance-evading drugs.

Five large conserved regions are identified in the enzyme structure derived from the drug-naive patients. These areas contain mainly important functional domains of the PR, such as the N and C termini, that are crucial for dimerization (1–9; 94–99), the active site (21–32), the flap (44–56), and the 78–88 region. Overall, these results have consistency with a first analysis conducted upon 20 different isolates [30], and other subsequent studies [17,18,22,31], as well as with the recent paper of Wu *et al.* [25]. Despite general agreement, we found slight differences in the rate of conservation at a few residues (e.g., 89, where a higher residue variability was observed in our cohort when compared to the results of the study by Wu *et al.* [25], 6.3% vs. $< 1\%$). The different mutational pattern of the drug-naive patients PR between studies can probably be related to the limited number of patients analysed in some reports [18,30,31], to the circulation of distinct viruses in different geographic areas, and to patients' characteristics.

In PI-experienced patients, several conserved residues clustered in small blocks, suggesting that participation of consecutive residues in structural domains is required for cooperative function and sustainability of enzyme activity. Amino acid conservation was almost complete in both N and C termini, but decreased dramatically in the other large areas, leading to shrunken minimal domains for the D25–D29 active-site, the top of the flap G49–G52, and the G78–P81 and G86–R87 turns. This conservation pattern has similarity to that reported by Wu *et al.* [25], and the few discrepancies at residues 11, 50, 79, and 83, where variability was in the range of 1–4%, can be attributed mainly to the different PI regimens within the two patient cohorts. Clearly, and not surprisingly, drug pressure influences the ability of PR to harbour mutations in the binding site (D30, V32, I47, I84, V82). These mutations play a role in increasing the K_i of different inhibitors used in HAART [11,32,40], whereas in the absence of drug pressure the residues lining the binding cavity are fully conserved. This agrees with the evidence that pharmacologically selected mutations around the PR active site are allowed if compensated by the presence of other mutations that tend to restore a good overall catalytic efficiency [41,42].

Overall, the invariant residues found *in vivo* are important for catalytic activity (D25–D29), structural stability of each monomer (L38 and Y59), substrate recognition (A28, I50, and P81) [43], and dimer stabilization [37,38]: L5 and T96–F99 (at the N and C termini, respectively), R8, T26–G27 and D29 (at the base of the active-site), G49–G51 (at the tip of the flap), R87, and T91. In this respect, we note the existence of a dual interaction (hydrophobic and hydrogen bond) between Y59 and L38 in each subunit that is likely to be important for monomer stabilization. Additionally, residues P1, I3, L5, and C95–F99, which have been shown to contribute close to 75% of the total free energy of dimerization [44], are found here to be invariant, with the only possible exception of C95, which was mutated in few PI-treated patients (2.7%). This result supports previous evidence, based on thermodynamic analysis, claiming that most of the dimerization energy is provided by the interface and that the isolated monomers are intrinsically unstable

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

[44]. Regarding the amino acid type present at invariant sites, it is noteworthy that, with the exception of G16 and the resistance-associated G48 and G73, all other glycines and all prolines are highly conserved. These data confirm a definitive structural role for these residues, which provide (to a larger extent than other amino acids) increased flexibility or rigidity, respectively. In addition, the overall trend of introducing conservative substitutions highlights the preference of several small changes (isolated or in combination) over a single dramatic substitution in the development of drug resistance.

Finally, it should be noted that many of the invariant residues detected in our work (L5, D25–D29, G40, G49, G52, V56, Y59, G68, V75–L76, G78, T80–P81, N83, G86–R87, L97) have been previously shown to be exquisitely sensitive to mutation by mutagenesis studies [45], thus confirming their key role and the necessity for their conservation. In contrast, the fact that some amino acids, E65, N83 and L90, which seem sensitive to mutation *in vitro* [45] and are found mutated *in vivo* by us and/or others, emphasizes the value of *in vivo* confirmation of previous *in vitro* observations, as well as the need to collect a large amount of *in vivo* data in order to be able to optimally define the minimal degree of conservation of this enzyme.

Some of the mutations observed in this study are the result of pharmacological pressure imposed by the drug regimens utilized. It can be argued that other regimens or new drugs might cause the appearance of mutations in PR regions found unchanged in our cohort of treated patients. This might be the case for invariant residues 8, 11, 28, 50, 79 and 91, which have been rarely found mutated within viral populations under different pharmacological regimens [25,46–51]. A recent report, however, suggests that the large majority of the new mutations described for the new drugs in current clinical development, such as tipranavir or atazanavir, typically develop in residues that have been previously associated to resistance to other PI [27]. Moreover, the importance of conserving this core enzyme structure is confirmed by the observation that selected changes in the minimal conserved structure under different selection pressures (e.g., I50V/L) are also quite detrimental for the enzymatic function and thus a high price would be paid in terms of viral fitness.

In conclusion, we present a three-dimensional model of HIV-PR conservation that identifies areas that are highly relevant for dimer stability, substrate recognition, and catalytic activity. Determining the biochemical and biophysical properties of enzymes with mutations in these invariant regions will clarify their effective role in the preservation of a stable and functional enzyme. Utilization of the definition of these minimal invariant areas should aid the design of inhibitors and could lead to novel PR-directed resistance-evading drugs.

Acknowledgements

We thank Sara Giannella, Maria Santoro, and Valentina Svicher and all the I.CO.N.A study group participants and members.

Sponsorship: Supported by grants from the National Institute for Infectious Diseases, Rome, the Italian National Institute of Health, the Ministry of University and Scientific Research, Current and Finalized Research of the Italian Ministry of Health and the European Community (QLK2-CT-2000-00291, and Descartes Award). The I.CO.N.A. network is supported by unrestricted educational grants from Glaxo-Smith-Kline, Italy.

References

1. Coffin JM. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science* 1995, **267**:483–489.
2. Perrin L, Kaiser L, Yerly S. **Travel and the spread of HIV-1 genetic variants.** *Lancet Infect Dis* 2003, **3**:22–27.
3. Wei X, Ghosh SK, Taylor ME, Johnson VA, Emini EA, Deutsch P, *et al.* **Viral dynamics in human immunodeficiency virus type 1 infection.** *Nature* 1995, **373**:117–122.
4. Condra JH, Schleif WA, Blahy OM, Gabryelski LJ, Graham DJ, Quintero JC, *et al.* **In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors.** *Nature* 1995, **374:**569–571.
5. Gonzales MJ, Wu TD, Taylor J, Belitskaya I, Kantor R, Israelski D, *et al.* **Extended spectrum of HIV-1 reverse transcriptase mutations in patients receiving multiple nucleoside analog inhibitors.** *AIDS* 2003, **17:**791–799.
6. Hertogs K, Bloor S, Kemp SD, Van den Eynde C, Alcorn TM, Pauwels R, *et al.* **Phenotypic and genotypic analysis of clinical HIV-1 isolates reveals extensive protease inhibitor cross-resistance: a survey of over 6000 samples.** *AIDS* 2000, **14**: 1203–1210.
7. Shafer RW, Stevenson D, Chan B. **Human immunodeficiency virus reverse transcriptase and protease sequence database.** *Nucl Acids Res* 1999, **27**:348–352.
8. Flexner CW. **HIV-protease inhibitors.** *New Engl J Med* 1998, **338**:1281–1292.
9. Huff JR. **HIV protease: a novel chemotherapeutic target for AIDS.** *J Med Chem* 1991, **34**:2305–2314.
10. Rodriguez-Barrios F, Gago F. **HIV protease inhibition: limited recent progress and advances in understanding current pitfalls.** *Curr Top Med Chem* 2004, **4**:671–686.
11. Wlodawer A, Erickson JW. **Structure-based inhibitors of HIV-1 protease.** *Annu Rev Biochem* 1993, **62**:543–585.
12. Alexander CS, Dong W, Chan K, Jahnke N, O'Shaughnessy MV, Mo T, *et al.* **HIV protease and reverse transcriptase variation and therapy outcome in antiretroviral-naive individuals from a large North American cohort.** *AIDS* 2001, **15**:601–607.
13. Boden D, Hurley A, Zhang L, Cao Y, Guo Y, Jones E, *et al.* **HIV-1 drug resistance in newly infected individuals.** *JAMA* 1999, **282**:1135–1141.
14. Frater AJ, Beardall A, Ariyoshi K, Churchill D, Galpin S, Clarke JR, *et al.* **Impact of baseline polymorphisms in RT and protease on outcome of highly active antiretroviral therapy in HIV-1-infected African patients.** *AIDS* 2001, **15**:1493–1502.
15. Gonzales MJ, Machekano RN, Shafer RW. **Human immu-nodeficiency virus type 1 reverse-transcriptase and protease subtypes: classification, amino acid mutation patterns, and prevalence in a**

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

northern California clinic-based population. *J Infect Dis* 2001, **184**:998–1006.
16. Grossman Z, Vardinon N, Chemtob D, Alkan ML, Bentwich Z, Burke M, *et al.* **Genotypic variation of HIV-1 reverse transcriptase and protease: comparative analysis of clade C and clade B.** *AIDS* 2001, **15**:1453–1460.
17. Kozal MJ, Shah N, Shen NP, Yang R, Fucini R, Merigan TC, *et al.* **Extensive polymorphisms observed in HIV-1 clade-B protease gene using high-density oligonucleotide arrays.** *Nat Med* 1996, **2**:753–759.
18. Lech WJ, Wang G, Yang YL, Chee Y, Dorman K, McCrae D, *et al.* **In vivo sequence diversity of the protease of human immunodeficiency virus type 1: presence of protease inhibitor-resistant variants in untreated subjects.** *J Virol* 1996, **70**:2038–2043.
19. Little SJ, Holte S, Routy JP, Daar ES, Markowitz M, Collier AC, *et al.* **Antiretroviral-drug resistance among patients recently infected with HIV.** *New Engl J Med* 2002, **347**:385–394.
20. Perno CF, Cozzi-Lepri A, Balotta C, Bertoli A, Violin M, Monno L, *et al.* **Low prevalence of primary mutations associated with drug resistance in antiviral-naive patients at therapy initiation.** *AIDS* 2002, **16**:619–624.
21. Perno CF, Cozzi-Lepri A, Balotta C, Forbici F, Violin M, Bertoli A, *et al.* **Secondary mutations in the protease region of human immunodeficiency virus and virologic failure in drug-naive patients treated with protease inhibitor-based therapy.** *J Infect Dis* 2001, **184**:983–991.
22. Pieniazek D, Rayfield M, Hu DJ, Nkengasong J, Wiktor SZ, Downing R, *et al.* **Protease sequences from HIV-1 group M subtypes A-H reveal distinct amino acid mutation patterns associated with protease resistance in protease inhibitor-naive individuals worldwide. HIV Variant Working Group.** *AIDS* 2000, **14**:1489–1495.
23. Ribeiro RM, Bonhoeffer S, Nowak MA. **The frequency of resistant mutant virus before antiviral therapy.** *AIDS* 1998, **12**:461–465.
24. Verbiest W, Brown S, Cohen C, Conant M, Henry K, Hunt S, *et al.* **Prevalence of HIV-1 drug resistance in antiretroviral-naive patients: a prospective study.** *AIDS* 2001, **15**:647–650.
25. Wu TD, Schiffer CA, Gonzales MJ, Taylor J, Kantor R, Chou S, *et al.* **Mutation patterns and structural correlates in human immunodeficiency virus type 1 protease following different protease inhibitor treatments.** *J Virol* 2003, **77**:4836–4847.
26. Yerly S, Kaiser L, Race E, Bru JP, Clavel F, Perrin L. **Transmission of antiretroviral-drug-resistant HIV-1 variants.** *Lancet* 1999, **354**:729–733.
27. Johnson VA, Brun-Vezinet F, Clotet B, Conway B, D'Aquila RT, Demeter LM, *et al.* **Drug resistance mutations in HIV-1.** *Top HIV Med* 2003, **11**:215–221.
28. Hirsch MS, Brun-Vezinet F, D'Aquila RT, Hammer SM, Johnson VA, Kuritzkes DR, *et al.* **Antiretroviral drug resistance testing in adult HIV-1 infection.** *JAMA* 2000, **283**:2417–2426.
29. Schinazi RF, Larder BA, Mellors J. **Mutations in retroviral genes associated with drug resistance: 2000-2001 update.** *Int Antiviral News* 2000, **8**:65–91.
30. Stanford HIV sequence database 2002: http://hivdb.Stanford.edu.
31. Fontenot G, Johnston K, Cohen JC, Gallaher WR, Robinson J, Luftig RB. **PCR amplification of HIV-1 proteinase sequences directly from lab isolates allows determination of five conserved domains.** *Virology* 1992, **190**:1–10.
32. Winslow DL, Stack S, King R, Scarnati H, Bincsik A, Otto MJ. **Limited sequence diversity of the HIV type 1 protease gene from clinical isolates and in vitro susceptibility to HIV protease inhibitors.** *AIDS Res Hum Retroviruses* 1995, **11**:107–113.
33. Wang W, Kollman PA. **Computational study of protein specifmanicity: the molecular basis of HIV-1 protease drug resistance.** *Proc Natl Acad Sci USA* 2001, **98**:14937–14942.
34. ClustalW: http://www.ebi.ac.uk.
35. Guex N, Peitsch MC. **Swiss-model and the Swiss-PdbViewer: an environment for comparative protein modelling.** *Electrophoresis* 1997, **18**:2714–2723.
36. Priestle JP, Fassler A, Rosel J, Tintelnot-Blomley M, Strop P, Grutter MG. **Comparative analysis of the X-ray structures of HIV-1 and HIV-2 proteases in complex with CGP 53820, a novel pseudosymmetric inhibitor.** *Structure* 1995, **3**:381–389.
37. Weber IT. **Comparison of the crystal structures and intersubunit interactions of human immunodeficiency and Rous sarcoma virus proteases.** *J Biol Chem* 1990, **265**:10492–10496.
38. Wlodawer A, Gustchina A. **Structural and biochemical studies of retroviral proteases.** *Biochim Biophys Acta* 2000, **1477**:16–34.
39. Wlodawer A, Miller M, Jaskolski M, Sathyanarayana BK, Baldwin E, Weber IT, *et al.* **Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease.** *Science* 1989, **245**:616–621.
40. Olsen DB, Stahlhut MW, Rutkowski CA, Schock HB, vanOlden AL, Kuo LC. **Non-active-site changes elicit broad-based cross-resistance of the HIV-1 protease to inhibitors.** *J Biol Chem* 1999, **274**:23699–23701.
41. Martinez-Picado J, Savara AV, Shi L, Sutton L, D'Aquila RT. **Fitness of human immunodeficiency virus type 1 protease inhibitor-selected single mutants.** *Virology* 2000, **275**:318–322.
42. Martinez-Picado J, Savara AV, Sutton L, D'Aquila RT. **Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1.** *J Virol* 1999, **73**:3744–3752.
43. Scott WR, Schiffer CA. **Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance.** *Struct Fold Des* 2000, **8**:1259–1265.
44. Todd MJ, Semo N, Freire E. **The structural stability of the HIV-1 protease.** *J Mol Biol* 1998, **283**:475–488.
45. Loeb DD, Swanstrom R, Everitt L, Manchester M, Stamper SE, Hutchison CA III. **Complete mutagenesis of the HIV-1 protease.** *Nature* 1989, **340**:397–400.
46. Carrillo A, Stewart KD, Sham HL, Norbeck DW, Kohlbrenner WE, Leonard JM, *et al.* **In vitro selection and characterization of human immunodeficiency virus type 1 variants with increased resistance to ABT-378, a novel protease inhibitor.** *J Virol* 1998, **72**:7532–7541.
47. Colonno RJ, Thiry A, Limoli K, Parkin N. **Activities of Atazanavir (BMS-232632) against a Large Panel of Human Immunodeficiency Virus Type 1 Clinical Isolates Resistant to One or More Approved Protease Inhibitors.** *Antimicrob Agents Chemother* 2003, **47**:1324–1333.
48. Ho DD, Toyoshima T, Mo H, Kempf DJ, Norbeck D, Chen CM, *et al.* **Characterization of human immunodeficiency virus type 1 variants with increased resistance to a C2-symmetric protease inhibitor.** *J Virol* 1994, **68**:2016–2020.
49. Maguire M, Shortino D, Klein A, Harris W, Manohitharajah V, Tisdale M, *et al.* **Emergence of resistance to protease inhibitor amprenavir in human immunodeficiency virus type 1-infected patients: selection of four alternative viral protease genotypes and influence of viral susceptibility to coadministered reverse transcriptase nucleoside inhibitors.** *Antimicrob Agents Chemother* 2002, **46**:731–738.
50. Maguire MF, Guinea R, Griffin P, Macmanus S, Elston RC, Wolfram J, *et al.* **Changes in human immunodeficiency virus type 1 Gag at positions L449 and P453 are linked to I50V protease mutants in vivo and cause reduction of sensitivity to amprenavir and improved viral fitness in vitro.** *J Virol* 2002, **76**: 7398–7406.
51. Yoshimura K, Kato R, Kavlick MF, Nguyen A, Maroun V, Maeda K, *et al.* **A potent human immunodeficiency virus type 1 protease inhibitor, UIC-94003 (TMC-126), and selection of a novel (A28S) mutation in the protease active-site.** *J Virol* 2002, **76**:1349–1358.

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.