Mutational Analysis of the Substrate Binding Pockets of the Rous Sarcoma Virus and Human Immunodeficiency Virus-1 Proteases*

Craig E. Cameron‡, Todd W. Ridky§, Sergey Shulenin, and Jonathan Leis¶

From Case Western Reserve University School of Medicine, Cleveland, Ohio 44106-4935

Irene T. Weber

From the Jefferson Cancer Institute, Thomas Jefferson University, Philadelphia, Pennsylvania 19107

Terry Copeland and Alexander Wlodawer

From the National Cancer Institute, National Institutes of Health, Frederick Cancer Research and Development Center, Advanced BioScience Laboratory-Basic Research Program, Frederick, Maryland 21702

Haim Burstein, Diane Bizub-Bender, and Anna Marie Skalka

From the Fox Chase Cancer Center, Institute for Cancer Research, Philadelphia, Pennsylvania 19111

Mutations, designed by analysis of the crystal structures of Rous sarcoma virus (RSV) and human immunodeficiency virus type 1 (HIV-1) protease (PR), were introduced into the substrate binding pocket of RSV PR. The mutations substituted nonconserved residues of RSV PR, located within 10 Å of the substrate, for those in structurally equivalent positions of HIV-1 PR. Changes in the activity of purified mutants were detected *in vitro* by following cleavage of synthetic peptides representing wild-type and modified RSV and HIV-1 *gag* and *pol* polyprotein cleavage sites. Substituting threonine for valine 104 (V104T), S107N, I44V, Q63M or deletion of residues 61–63 produced enzymes that were 2.5–7-fold more active than the wild type RSV PR. Substituting I42D, M73V, and A100L produced enzymes with lower activity, whereas a mutant that included both M73V *and* A100L was as active as wild type. Several substitutions altered the specificity for substrate. These include I42D and I44V, which contribute to the S2 and S2′ subsites. These proteins exhibited HIV-1 PR specificity for P2- or P2′-modified peptide substrates but unchanged specificity with P4-, P3-, P1-, P1′-, and P3′-modified substrates. Changes in specificity in the S4 subsite were detected by deletion of residues 61–63. These results confirm the hypothesis that the subsites of the substrate binding pocket of the retroviral protease are capable of acting independently in the selection of substrate amino acids.

The retroviral protease (PR)[1] is responsible for the post-translational processing of the *gag* and *pol* polyprotein precur-

sors of nascent virus particles (1). This process, termed maturation, is necessary for the production of infectious virions (1). Loss of PR activity either by the presence of inhibitors or by mutations results in the production of immature, noninfectious particles (2–4). As a result, PR is a target for the design of antiviral agents.

PRs from different retroviruses efficiently recognize and process sequences present in their cognate *gag* and *pol* polyproteins (5). Yet there is no simple consensus sequence when one compares the sequences known to be specifically and efficiently processed by a particular PR. Additionally, heterologous activity, that is, activity of PR from one retrovirus on the *gag* and *pol* polyprotein cleavage sites of another retrovirus, has been reported (6). The PRs possess intricately designed substrate binding pockets that contain at least seven subsites (S4-S3′) which interact with 7 amino acids of the substrate (P4-P3′) (7, 8). Although the general topologies of the substrate binding pockets of different PRs are virtually identical, subtle differences in structure and amino acid composition do exist (9). These are likely responsible for the differences in activities observed among the various proteases. For instance, the RSV PR is 10–20-fold less active catalytically than HIV-1 PR (1, 6). Also, the RSV PR is more stringent in the selection of substrates than HIV-1 PR (6, 9); HIV-1 PR is capable of specifically and efficiently processing peptides representing the RSV *gag* and *pol* polyprotein cleavage sites, whereas the converse is not true.

We have used site-directed mutagenesis to investigate the significance of nonconserved amino acid residues of the PR substrate binding pockets. Using crystallographic information available for the RSV and HIV-1 enzymes (9), we identified 11 amino acid residues of the RSV PR which are within 10 Å of the substrate and are not identical in the HIV-1 PR. These residues are: Ser-38, Ile-42, Ile-44, Gln-63, His-65, Met-73, Ala-100, Val-104, Arg-105, Gly-106, and Ser-107. The residues in the structurally analogous positions of the HIV-1 PR are *Thr-26, Asp-30, Val-32, Met-46, Gly-48, Val-56, Leu-76, Thr-80, Pro-81, Val-82,* and *Asn-83,* respectively. (For clarity, amino acids derived from HIV-1 PR will be placed in *italics* throughout the text and the one letter notation for amino acids will be used to refer to substitutions and mutations.)

In a previous report, we described results from substitutions that affected the S1 and S1′ subsites of the RSV PR substrate

TABLE I
RSV PR mutations

Mutation[a]	Oligodeoxynucleotide[b]	Restriction enzyme[c]
I42D	CGCTGTTGGACTCTGGCGCCGACGACACTATTATTTCAGAGG3'	KasI
I44V	TGGAGCGGACATCACTGTGATATCAGAGGAGGATTGGC3'	EcoRV
Q63M	GGAGGCCGCGAACCCGATGATTCATGGGATAGGAGGGG3'	NcoI[d]
H65G	AGCAGATCTCTCACCTACCAAACAATG3'	BstXI
	TGGGAATTCCCCCTCCTATCCCACCGATCTGGGGGTTCGCG GCCTCCA3'	
M73V	AGGAGGGGGAATTCCCGTTCGAAAATCTCGTGACA3'	SfuI
A100L	CCCTGCTCCTCTTCCCACTAGTAGCTATGGTTAGA3'	SpeI
M73V, A100L	Same as above	SfuI/SpeI
V104T	CCCCGCAGTAGCTATGACGCGTGGGAGTATCCTAGGAA3'	MluI
S107N	GC TATGGTTAGAGGGAATATTCTAGGAAGAGATTGTC3'	SspI
N61P, P62L, Q63M	AGTGATGGAGGCCGCGCCCCTCATGATTCATGGGATAG3'	BspHI
Δ61–63	AGTGATGGAGGCCGCGATTCATGGGATAGGAGGGGGA3'	NcoI[d]
R105P, G106V	CCGCAGTAGCTATGGTACCAGTGAGTATCCTAGGAAGA3'	KpnI
R105P, G106V, S107N	AGCTATGGTACCAGTGAATATTCTAGGAAGAGATTGTC3'	KpnI/SspI
V104T, R105P, G106V, S107N	TCCCCGCAGTAGCTATGACACCGGTGAATATTCTAGG AAGA3'	AgeI/SspI

[a] Mutations are identified by amino acid and position number in the RSV PR followed by the substituted amino acid (italicized). Amino acids are denoted by the single letter code. Deletion mutations are indicated by Δ followed by the position number of the first and last deleted amino acid. Multiple amino acid substitutions are indicated by the commas.

[b] Oligodeoxynucleotides used to place mutations into the pPR plasmid by site-directed mutagenesis.

[c] New restriction enzyme sites (underlined) were introduced to facilitate identification of mutated clones.

[d] Indicates the loss of the original NcoI restriction site.

binding pocket (H65G, R105P, and G106V) and produced mutant PRs with expanded substrate specificity (10, 11). A mutant RSV PR containing all three substitutions was capable of recognizing a HIV-1 *pol* polyprotein cleavage site 50-fold more efficiently than the wild type RSV PR. However, the catalytic efficiency on other HIV-1 cleavage site substrates was unchanged. Subsequent analysis of the activity of this mutant on a library of peptide substrates containing single amino acid substitutions in the P4-P3' positions showed that the HIV-1 PR-like behavior of this mutant PR could be observed exclusively with P1- and P1'-modified peptide substrates (10). This suggested that subsites of the substrate binding pocket acted independently in the selection of substrate (9). In this report, we describe the expression, purification, and characterization of mutant RSV PRs with single amino acid substitutions in the other subsites of the substrate binding pocket. The results of this analysis confirm our earlier work and identify specificity determinants in subsites S2, S2', and S4. In addition, we have identified amino acid residues within the substrate binding pocket which contribute to the differences in catalytic efficiency of the RSV and HIV-1 PRs.

EXPERIMENTAL PROCEDURES

Bacterial Cells and DNA Constructs—Construction of the wild type RSV PR and NC-PR expression vectors was described previously (10). *Escherichia coli* MC 1061 cells transformed with the temperature-sensitive λ-cI repressor plasmid, pRK248cIts, were used for the expression of the mutant PR proteins.

Mutagenesis—Standard site-directed mutagenesis and oligodeoxynucleotide purification procedures were used to introduce amino acid changes (substitutions and/or deletions) into the protease coding region of the plasmids pPR or pNC-PR (10). The PR mutations, oligodeoxynucleotides used in the mutagenesis reaction, and the diagnostic restriction enzyme sites are shown in Table I. Each construct was first verified by digesting the DNA with the respective diagnostic restriction enzyme and later by sequencing.

To construct the PR mutants I42D; I44V; Q63M; M73V; M73V,A100L; V104T; S107N; N61P,P62L,Q63M; Δ(61–63); V104T, R105P,G106V,S107N the appropriate purified oligodeoxynucleotide was annealed to linearized (*Pvu*I) and gapped (*Bgl*II-*Bss*HII) DNA from wild type PR. To construct the PR R105P,G106V,S107N mutant, the oligodeoxynucleotide was annealed to purified linear (*Pvu*I) and gapped (*Bgl*II-*Bss*HII) DNA from PR R105P,G106V. NC-PR R105P,G106V was

constructed by annealing the purified oligodeoxynucleotide to purified and linearized (*Pvu*I) and gapped (*Bgl*II-*Bss*HII) DNA from the wild type NC-PR construct.

The polymerase chain reaction was used to construct PR H65G. The designated purified oligodeoxynucleotides were annealed to wild type PR DNA, and VENT DNA polymerase was used to amplifiy the DNA. The amplified DNA fragment was then digested with *Bgl*II and *Eco*RI and the DNA purified. The 0.48-kb *Bgl*II-*Eco*RI fragment was then ligated to the purified 2.7-kb DNA fragment from wild type PR to yield PR H65G.

PR R105P,G106V was produced by ligating the purified 0.48-kb *Bgl*II-*Eco*RI fragment from PR to the 2.7-kb DNA fragment from NC-PR R105P,G106V. PR A100L was produced by ligating the 0.55-kb *Mse*I-*Acc*I fragment from the aberrant PR M73V,A100L construct to the 2.5-kb dephosphorylated *Pfl*MI-*Acc*I fragment and the 96-base pair *Pfl*MI-*Mse*I fragments from wild type PR.

Preparation of Retroviral Proteases—Wild type and mutant proteins were prepared from *E. coli* grown in LB medium at pH 6.8 and induced for expression of RSV PR as described previously (11). Cells harvested and washed by centrifugation were suspended in 10 mM Tris·HCl, pH 8.0, in 1/10 the culture volume and disrupted with EDTA (10 mM) and lysozyme (1 mg/ml) followed by treatment with Triton X-100 (1%). The pellet was collected and washed by sequential suspension and centrifugation (four times), first with 25% sucrose, 1% Triton X-100 and then with sucrose/Triton X-100 containing 4 M and then 6 M urea. The inclusion bodies were solubilized in 8 M urea and 150 mM 2-mercaptoethanol, and the protease was refolded by dialysis (0.5–1 mg of protein/ml) at 23 °C against a decreasing concentration of urea (4, 2, and 1 M for 2 h each). The final dialysis (against 25 mM sodium phosphate, pH 7.5, 150 mM NaCl, and 10 mM 2-mercaptoethanol) was overnight. The clarified soluble fraction containing active PR was stored at 4 °C. The preparations were at least 95% pure as determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis with Coomassie staining (12). Protein concentrations were determined using the Bio-Rad protein assay kit and bovine serum albumin as a standard. In some experiments, wild type and mutant PR subunits were mixed in 8 M urea and renatured together.

AMV PR, purified from virus as described (13), was obtained from Molecular Genetic Resources, Tampa, FL. HIV-1 PR expressed in and purified from bacteria was a generous gift of Dr. Joe Giam, Case Western Reserve University.

Peptides—The peptides were synthesized and purified as described previously (6). Peptides were solubilized in 1 mM 2-mercaptoethanol, and their concentrations were determined by quantitative amino acid composition analysis. Specificity of cleavage was established by direct amino-terminal analysis of product peptides (6).

Assay of PR Activity—The reaction mixture contained 100 mM sodium phosphate, pH 5.9, 2.4 M sodium chloride, 100 μM peptide, and

FIG. 1. **Schematic representation of the AMV/RSV NC-PR substrate, PAVSLAMT from P4 to P4' in the S4 to S4' subsites of PR.** The relative size of each subsite is illustrated by the area enclosed by the *curved line* around each substrate side chain. Protease residues forming the subsites are shown for those that differ between the AMV/RSV and HIV-1 PRs. The AMV/RSV PR residue is shown *outside* of the *parentheses*, and the HIV-1 PR residue is shown in the *parentheses*. Most of the residues contribute to more than one adjacent subsite, and this is indicated by the position of the label.
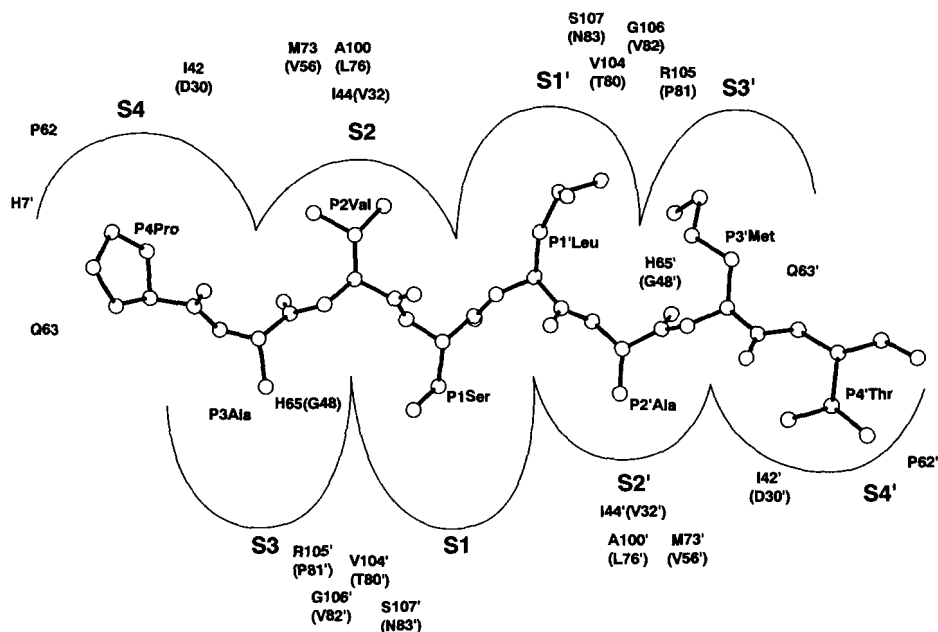
50–500 nM active PR. Reaction volumes were 25 μl. Reactions were initiated by the addition of PR, incubated for various periods of time, and stopped by the addition of 300 μl of 0.5 M sodium borate, pH 8.5. Twenty μl of 0.05% (w/v) fluorescamine was then added. HIV-1 PR was never incubated more than 8 min because of its instability, presumably a result of autodegradation. After reaction with fluorescamine, the relative fluorescence was determined on a Perkin-Elmer LS-50B spectrofluorometer using an excitation wavelength of 386 nm and an emission wavelength of 477 nm. Excitation and emission slit widths were 5 and 10 nm, respectively. Relative fluorescence intensity was converted to nmol of product using a standard curve described by the following equation: nmol of product = relative fluorescence intensity/313. The standard curve was obtained using a hexapeptide with a free amino terminus (11). The concentration of this peptide was determined by amino acid composition analysis. The peptides used in this study were designed with prolines at their amino termini so that the relative fluorescence intensity represents only the newly formed amino termini produced as a result of proteolytic cleavage. Each activity measurement reported represents the mean of at least three independent experiments. In each case, the standard error for all experiments did not exceed 20% of the value reported.

Determination of Efficiency of Refolding PR—PR, 50–500 nM, (E_t) was incubated in a reaction mixture containing 100 μM NC-PR and between 0 and 500 nM inhibitor peptide (PPCV-PheSta-AMTM) as described above. The PPCV-PheSta-AMTM peptide is, in our hands, a 5–20 nM inhibitor of PR, depending upon the enzyme used. The velocity (V) of each reaction was plotted as a function of the concentration of inhibitor. The data were then fit to the following equation described by Cha (14) using the program GraFit (Erithacus Software Limited).

$$V = \frac{k}{F[E_t]} (F[E_t] - 0.5(F[E_t] + [I_t] + K_i) +$$
$$0.5((F[E_t] + [I_t] + K_i)^2 - 4F[E_t][I_t]))^{1/2}$$

(Eq. 1)

where k is the initial velocity of the reaction in the absence of inhibitor, F is the fraction folded, I_t is the inhibitor concentration, and K_i is the inhibition constant. Both F and K_i were allowed to vary.

Molecular Modeling—The crystal structures of the native RSV PR (15) and HIV-1 PR, as well as the complexes of the latter with inhibitors (16–20) were superimposed using C_α atoms and examined on an Evans and Sutherland ESV10 computer graphics system using the program FRODO (21). The residues forming the subsites were determined directly for the HIV-1 PR complexed with inhibitors, and the corresponding residues in RSV PR were obtained from the structural and sequence alignments (10). The NC-PR substrate was modeled as described previously (10), and different amino acid side chains were substituted at positions P4 to P3' of the substrate and at the mutations in RSV PR to determine the structural basis for the kinetic data.

RESULTS

Rationale for Mutations in the RSV PR—The ability to design RSV PRs with partial HIV-1 PR specificity by changing only nonconserved amino acids in the S1 and S1' subsites (10, 11) suggested that a complete change to HIV-1 PR specificity might be achieved by identification and simultaneous replacement of all key amino acid residues in the five additional RSV PR subsites. To identify these residues, a detailed structural and kinetic analysis of subsites of RSV and HIV-1 PRs was conducted (9). A list of potential amino acids that could influence substrate specificity is provided in Table I and includes His-65, Arg-105, Gly-106, Ile-42, Ile-44, Gln-63, Met-73, Ala-100, Val-104, and Ser-107. The subsites to which each residue has been assigned are indicated in Fig. 1. In each case, the amino acid found in the structurally analogous position of HIV-1 PR was placed into RSV PR by site-directed mutagenesis of the appropriate expression plasmid. Several mutant PRs were constructed which contained multiple substitutions. Also presented in Table I is the description of a mutation that deletes residues 61–63 from RSV PR. These residues are part of the long RSV PR flaps, and their removal is predicted to produce flaps of the same length as those of HIV-1 PR (10). In addition, our modeling suggested that these "extra" residues of RSV PR could potentially influence substrate interactions in the S4, S3, and S3' subsites.

Effect of Amino Acid Substitutions on the Activity of Mutant PRs—The activity of each RSV mutant with the wild type RSV NC-PR peptide substrate is presented in Table II. The activity data have been corrected for efficiency of protein refolding by active site titration of the enzymes using a nM peptide inhibitor as described under "Experimental Procedures." With the exception of the I42D mutant, all of the recombinant PRs refolded with similar efficiencies, although the degree of folding varied between 18 to 30% from preparation to preparation. The I42D protein tended to form a precipitate upon refolding. Many of the amino acid substitutions produced enzymes with specific activities greater than wild type RSV PR (Table II). Exceptions were the I42D, M73V, and A100L. Structural data suggested that the side chains of Met-73 and Ala-100 (Val-56 and Leu-76 in HIV-1 PR) could form intramolecular van der Waal's interactions. If this interaction is important for catalytic activity,

TABLE II

Activity of RSV proteases with mutations in the substrate binding pocket

PR mutation	Activity[a]
	min^{-1}
Wild type	13.0
I42D	8.0
I44V	27.4
Q63M	58.2
M73V	4.0
A100L	1.1
M73V, A100L	11.4
V104T	29.8
S107N	49.0
N61P,P62L,Q63M	56.4
ΔN61–Q63	31.8
R105P,G106V	12.7
R105P,G106V,S107N	95.9
V104T,R105P,G106V,S107N	64.5

[a] Activity was determined as described under "Experimental Procedures" using mutated RSV PRs as indicated and the RSV NC-PR peptide substrate. See Table I for explanation of italics and commas.

then substitution of one of these amino acids in the absence of the other is expected to produce PRs with low activity. This appears to be the case, since a double mutant containing both M73V and A100L substitutions is almost as active as wild type (Table II).

Of the mutant RSV PRs tested, those with substitutions at 63 and 107 were the most active; each resulted in at least a 4-fold increase in specific activity. The triple mutant N61P,P62L,Q63M also produced a more active enzyme. However, its specific activity was no greater than that observed with the Q63M substitution alone. This suggests that the increased activity of the triple mutant PR may be attributed primarily to the substitution at position 63. Increased activity was also observed when substitutions were placed at positions 44 and 104 and when residues 61–63 were deleted (see Table II).

Substitutions at residues 105 and 106 had previously been shown not to change the activity of the mutant PR (10). However, when the amino acid substitution at 107 was combined with those at 105 and 106, a 7-fold increase in specific activity was observed. However, combination of the substitutions at 104, 105, 106, and 107 produced an enzyme that was only 4-fold more active than wild type. Thus, the effects of single mutations on the rate of cleavage are not necessarily additive.

Activity of Mutant RSV PRs on Modified NC-PR Peptide Substrates—Each of the above mutants was examined for changes in substrate preference using the library of NC-PR peptides with amino acid substitutions in the P4 to P3' substrate positions. This set of peptides detects differences in the specificities of HIV-1 and RSV PRs (9, 10), and those analyzed here revealed large differences in activity in the two PRs. Among the mutations that increased the specific activity of the RSV PR, substitutions at positions 104 and 107 did not cause detectable changes in substrate preference. For instance, the V104T or the S107N mutation is predicted to influence the S1 and S1' subsites. Yet, there was no detectable change in activity of these mutants toward P1-substituted NC-PR peptide substrates as might be expected for increased HIV-1-like behavior (Fig. 2, E and F). This is in contrast to observations with the R105P,G106V substitutions using the same modified NC-PR substrates (10). When the V104T or S107N mutant was tested with NC-PR substrates that contained amino acid changes in other peptide positions, there were no new substrate preferences noted (Fig. 2, A and D).

In contrast to enzyme subsite substitutions at residues 104 and 107, substitution of residues 42 and 44 does induce changes in substrate preference. Both residues have been assigned to

the S2 and S2' subsites (9), and residue 42 is also near the S4 subsite (see Fig. 1). We tested each of these RSV mutants with substrates that change both valine in P2 (Fig. 2C) and alanine in P2' to leucine, a hydrophobic amino acid with a larger side chain (Fig. 2D). With the leucine P2 NC-PR peptide, both the 42 and the 44 enzyme substitution exhibited HIV-1-like specificity. However, with the leucine P2' NC-PR peptide, only the I44V substitution behaved like HIV-1 PR. Similar HIV-1-like specificity was observed with the NC-PR peptide substrate containing serine and glycine in P2 and alanine in P2' (data not shown). Although changes in specificity were detected with peptides with changes in positions P2 and P2', none was observed with changes at P3', P1', and P3 (data not shown), P1 (Fig. 2, E and F), or P4 (Fig. 2, A and B).

The I42D mutant showed increased activity toward a peptide containing histidine in the P4 position (Fig. 2A). This increase was not observed when the charge or size of the side chain at this position was changed by insertion of asparagine (Fig. 2B) or glycine (data not shown). This suggests that the increase observed with histidine is the result of a local ionic interaction between Asp-42 in the PR and histidine in the substrate.

Of the other mutants tested, only the Δ61–63 deletion mutant appeared to induce HIV-1-like preference for substrate amino acids in the P4 position of the NC-PR peptide (see Fig. 2, A and B). The mutant that contains N61P,P62L,Q63M did not exhibit altered specificity even though the specific activity of this enzyme relative to wild type was increased. Thus, the removal of residues 61–63 appears to be the important factor in changing substrate preference in the S4 enzyme subsite. As is the case for mutations at 42 and 44 and the S2 and S2' subsites, the Δ61–63 mutation appears to affect only the S4 subsite. Changes in substrate preference were not observed with NC-PR peptides altered at P2 or P2' (Fig. 2, B and C) or in the other positions (data not shown).

Activity of PR Mutants on Substrates Representing HIV Cleavage Sites—A series of peptide substrates that represent five of the nine cleavage sites in HIV-1 *gag* and *pol* polyproteins were used as substrates for the mutant RSV enzymes (Table III). The I44V mutant had a significantly increased rate of cleavage of the HIV-1 CA-NCb (Fig. 3B) and RT-IN (Fig. 3E) peptide substrates. The I42D mutant had increased activity toward the RT-IN (Fig. 3E) and NC-p6b (Fig. 3B) substrates. In the latter case, the HIV-1 NC-p6b substrate was cleaved at a rate comparable to that observed with the wild type HIV-1 PR. The Δ61–63 mutant increased the PR cleavage of the CA-NCb peptide substrate (Fig. 2B), whereas the double mutant M73V,A100L showed increased activity on the RT-IN substrate (Fig. 3D). The other mutants listed in Table II exhibited only marginal changes in activity toward these HIV-1 substrates (Fig. 3). In every case in which there was significant activity, the site of cleavage was established by amino-terminal analysis of product peptides and shown to be the same as with HIV-1 PR. Several mutants, including the I44V, Δ61–63, and S107N, showed small increases in activity toward the CA-NCa and/or the NC-p6a peptides (Fig. 2, A and C). None of these mutations affected cleavage of RSV peptide substrates (data not shown). This is consistent with the fact that the HIV-1 PR can cleave eight of the nine RSV peptide substrates (6).

DISCUSSION

Cleavage site specificities of the RSV and HIV-1 PRs are very broad. In their respective *gag* and *gag-pol* polyproteins, each efficiently cleaves *in vivo* nine different sequences (1). To understand the molecular basis for this specificity, we have analyzed each of the RSV and HIV-1 PR substrate binding pockets by substituting structurally analogous amino acids of HIV-1 PR into the RSV PR. This has allowed us to identify RSV PR
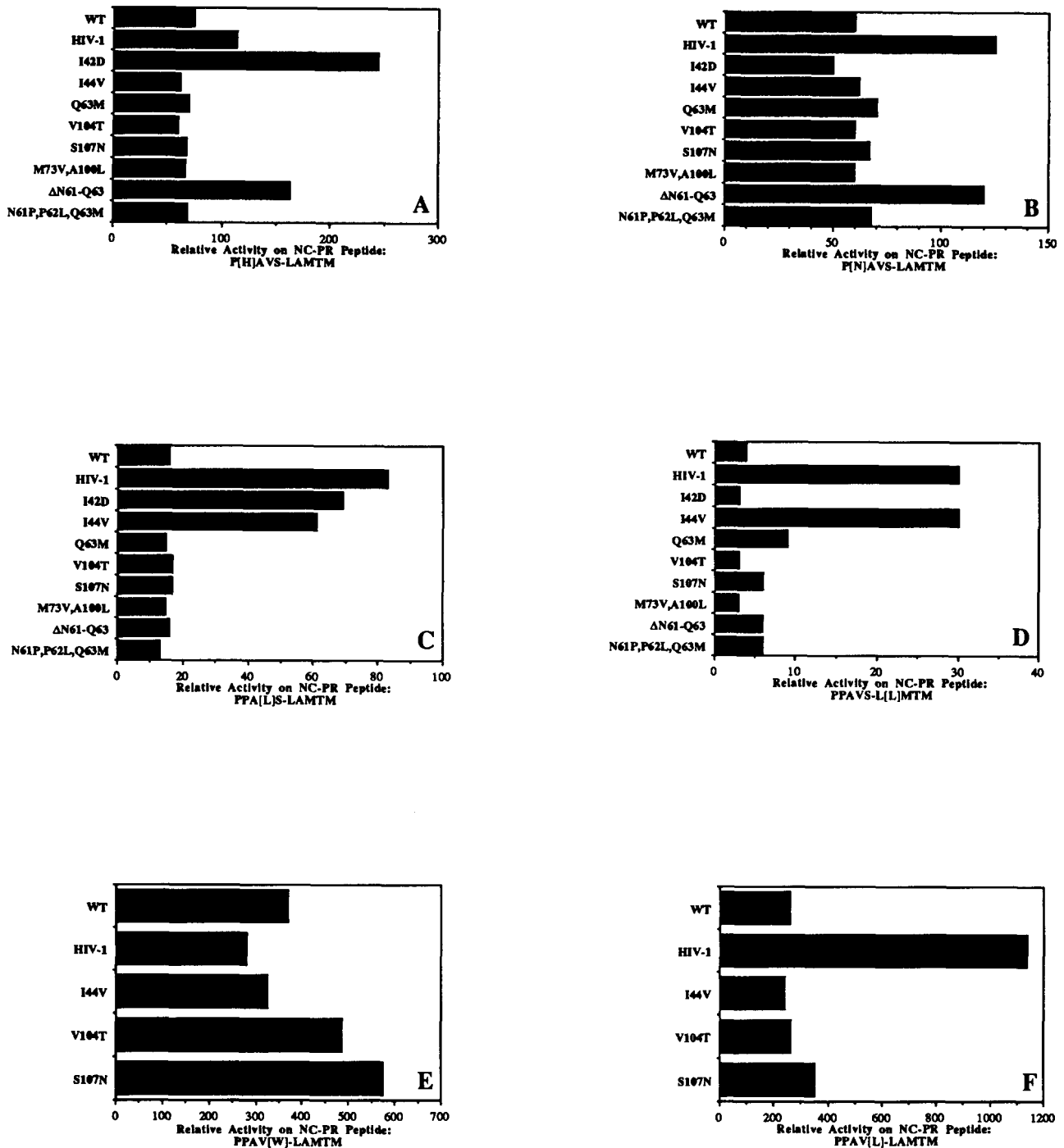
FIG. 2. **Protease activities with NC-PR peptide substrates substituted at the P4, P2, P1, and P2′ positions.** The NC-PR substrate (PPAVS-LAMTMRR) was modified with single amino acid substitutions in the positions indicated, and the activity with purified AMV PR, HIV-1 PR, and RSV mutant PRs (as described in Table II) was assayed as described under "Experimental Procedures." The data are expressed as activity with modified substrates relative to that observed with the wild type (*WT*) NC-PR peptide substrate. The following modified NC-PR peptide substrates were used: P[H]AVS-VAMTMRR (*panel A*), P[N]AVS-VAMTMRR (*panel B*), PPA[L]S-VAMTMRR (*panel C*), PPAVS-L[L]MTMRR (*panel D*), PPSV[W]-LAMTMRR (*panel E*), and PPAV[L]-LAMTMRR (*panel F*). The *dashed line* indicates the protease cleavage site. The *bracketed letter* denotes the amino acid that has been changed from the wild type sequence. The *nonbold letters* are amino acids not present in the polyproteins.

residues 42, 44, 65, 73, 100, 105, and 106 as those that dictate differences in specificity relative to HIV-1 PR (see Fig. 1). We have found that the deletion of residues 61–63 also influences substrate selection in the S4 subsite. By analogy, the respective structurally related residues, *30, 32, 48, 56, 76, 81,* and *82,* in HIV-1 PR must play an important role in substrate recognition. This has led to the prediction that these amino acids would be those that rapidly mutate to form resistant enzymes *in vivo*

when peptide-like inhibitors are employed as antivirals (9). This prediction has recently been verified by the isolation of drug-resistant variants with mutations at *Val-82* (22), *Val-32,* and *Gly-48.*[2]

The combined structural and molecular genetic approaches used in these studies have confirmed that, to a large extent,

---

[2] R. Swanstrom and J. Kay, personal communication.

TABLE III
Peptides used as PR substrates

Virus and location of cleavage site	Peptide sequence[a]
RSV	
gag NC-PR	PPAVS-LAMTMRR
HIV-1	
gag CA-NCa	PARVL-AEAMRR
gag CA-NCb	PATIM-MQRERR
gag NC-p6a	PGNF-LQSRR
gag NC-p6b	PRQAN-FLGKRR
pol RT-IN	PRKIL-FLDGRR

[a] Hyphens indicate the scissile bond. In addition to the P4-P3' amino acids assumed to interact with the PR, amino-terminal prolines were added to avoid reactivity with fluorescamine and carboxyl-terminal arginines to improve solubility.

individual enzyme subsites act independently in accommodating an amino acid side chain from a corresponding substrate position. As has been suggested earlier (9, 11, 25), the overall choice of substrate is dictated by the sum of all of the enzyme subsites. Thus, the relative independence of subsites provides a mechanism for accommodation of the wide variety of sequences present in retroviral polyprotein cleavage sites. Despite this independence, most of these critical residues in the PRs contribute to more than one subsite, and substrates with large or bulky side chains show effects that depend upon adjacent residues. For example, Tozser *et al.* (23) have described steric effects between residues in the P1' and P2 peptide substrate positions for the HIV-1 PR. For an analysis of the characteristics of each enzyme subsite, see Cameron *et al.* (9).

The nature of the interactions between amino acids in the substrate binding pockets and those in the substrate appears to involve mainly side chain van der Waal's forces. This is best illustrated with the substitution of I44V in the RSV S2 subsite (see Fig. 4). The NC-PR peptide substrate contains valine in the P2 position whose side chain is predicted to form an intermolecular van der Waal's interaction with Ile-44 of PR. If the enzyme's isoleucine is changed to *valine*, as in HIV-1 PR, the potential to form optimized van der Waal's interactions is lost, resulting in a significant decrease in activity. Wild type activity can be restored by substituting leucine for valine, presumably by adding the lost methylene group, critical for van der Waal's interactions, to the amino acid side chain in P2 position of substrate (Fig. 4).

Other examples of enhanced hydrophobic interactions with amino acid side chains in the substrate and enzyme can be found with RSV PR residues 73 and 100, and with deletion of residues 61–63. The former are present in the S2 and S2' subsites, and the latter are unique to the RSV PR and form part of the S4 subsite lying near the enzyme surface. The extra amino acids, 61–63, can explain why more nonpolar residues can be accommodated in the P4 position of substrate by the RSV PR than by the HIV-1 PR (9).

Although optimization of van der Waal's interactions appears to be extremely important for substrate selection in all of the subsites, ionic or H bond interactions are also possible at some positions. This is illustrated by the I42D mutant, which shows a preference for a basic residue in the P4 position of the NC-PR substrate (Fig. 3).

In exchanging amino acids from comparable structural positions of HIV-1 and RSV PRs, we have also identified residues that influence the rate of catalysis of the enzyme. These include both amino acids that influence the selection of substrate, such as Ile-42 and Ile-44, and deletion of residues 61–63, as well as those that do not, such as Val-104 and Ser-107. Val-104 and Ser-107 are involved in interactions that maintain the structural integrity of the PR catalytic site. In HIV-1 PR, *Thr-80* is the residue analogous to Val-104 in RSV PR. Structural anal-

ysis shows that *Thr-80* forms a hydrogen bond interaction with the carbonyl oxygen of *Val-82*. Thus the V104T mutant is predicted to introduce a similar interaction between the *threonine* hydroxyl and the carbonyl oxygen of Gly-106. This is predicted to increase the stability of the second ψ loop structure in the RSV PR which could lead to a higher catalytic rate. A similar argument can be made for the S107N mutation, in that *Asn-83* in HIV-1 PR can form hydrogen bond interactions with the amide of *Glu-34* and the carbonyl oxygen of *Glu-21*. The effect on the rate caused by mutation and deletion of residues 61–63 is more difficult to interpret since this region is not visible in the crystal structure of the RSV PR because of disorder. However, these residues were modeled as an insertion in the flap of RSV PR compared with that in HIV PR (10). Thus, their removal is expected to produce RSV flaps of the same length as those in HIV-1 PR. The increased rates of catalysis for the mutations that delete or alter these residues in RSV PR suggest that the shorter flaps of HIV PR are more favorable for product release. In general, we have found that mutations that affect the rate of cleavage are those that alter PR structural architecture, whereas those that influence substrate selectivity alter van der Waal's interactions between the side chains of amino acids in the substrate and enzyme.

The fact that HIV-1 PR has a more rigid active site structure than the RSV PR suggests that the HIV-1 PR may be less dependent upon substrate side chain interactions for activity. For instance, the HIV-1 PR tolerates glycine substituted in each of the NC-PR substrate positions, whereas RSV PR does not (9). This may also explain why the HIV-1 PR has a broader substrate specificity than the RSV PR and can cleave heterologous RSV polyprotein substrates (6).

The notion that the specificity is determined by the sum of contributions from individual subsites implies that within a given substrate not all amino acid positions will contribute equally to the interaction with the retroviral PRs. This may explain variations in the activity of the RSV PR mutants with the HIV-1 peptide substrates (Fig. 3). The substitution I44V results in a considerable increase in activity toward two of the HIV-1 polyprotein substrates, RT-IN and CA-NCb, but only modest increases in activity toward the three other substrates tested. In a similar fashion, the I42D mutant showed increases only with the NC-p6b and RT-IN substrates and the Δ61–63 mutant only with the CA-NCb peptide. The combination mutant M73V,A100L also showed activity toward the HIV-1 CA-NCb peptide substrate. This latter result indicates that these 2 residues are also important for substrate selection, probably because of their contributions to the S2 and S2' enzyme subsites. However, this assignment is based only on structural data since the mutant has not been analyzed with the modified NC-PR peptides.

A closer examination of the amino acid sequences of the HIV-1 cleavage peptides (see Table III) provides an explanation for the observed enzyme preferences. The RSV I44V mutant acts more effectively on the two HIV-1 substrates that contain isoleucine in the P2 position. Thus, the enhancement in activity can be related to improved van der Waal's interactions between isoleucine in P2 and *Val-44* in the PR, as illustrated in Fig. 4. The RSV I42D prefers HIV-1 substrates that contain basic residues in the P4 position. This is consistent with the data in Fig. 2D, which show increased activity of the I42D mutant toward a NC-PR peptide substrate containing histidine in the P4 position. The presence of histidine at P4 of the RSV NC-PR peptide improves the binding affinity for enzyme by about 2-fold (9). The activity of the RSV I42D mutant toward the NC-p6b substrate is equal or better than that seen with the wild type HIV-1 PR. This is consistent with the fact the NC-p6b substrate contains arginine in P4 and lysine in P4', yielding potential ionic
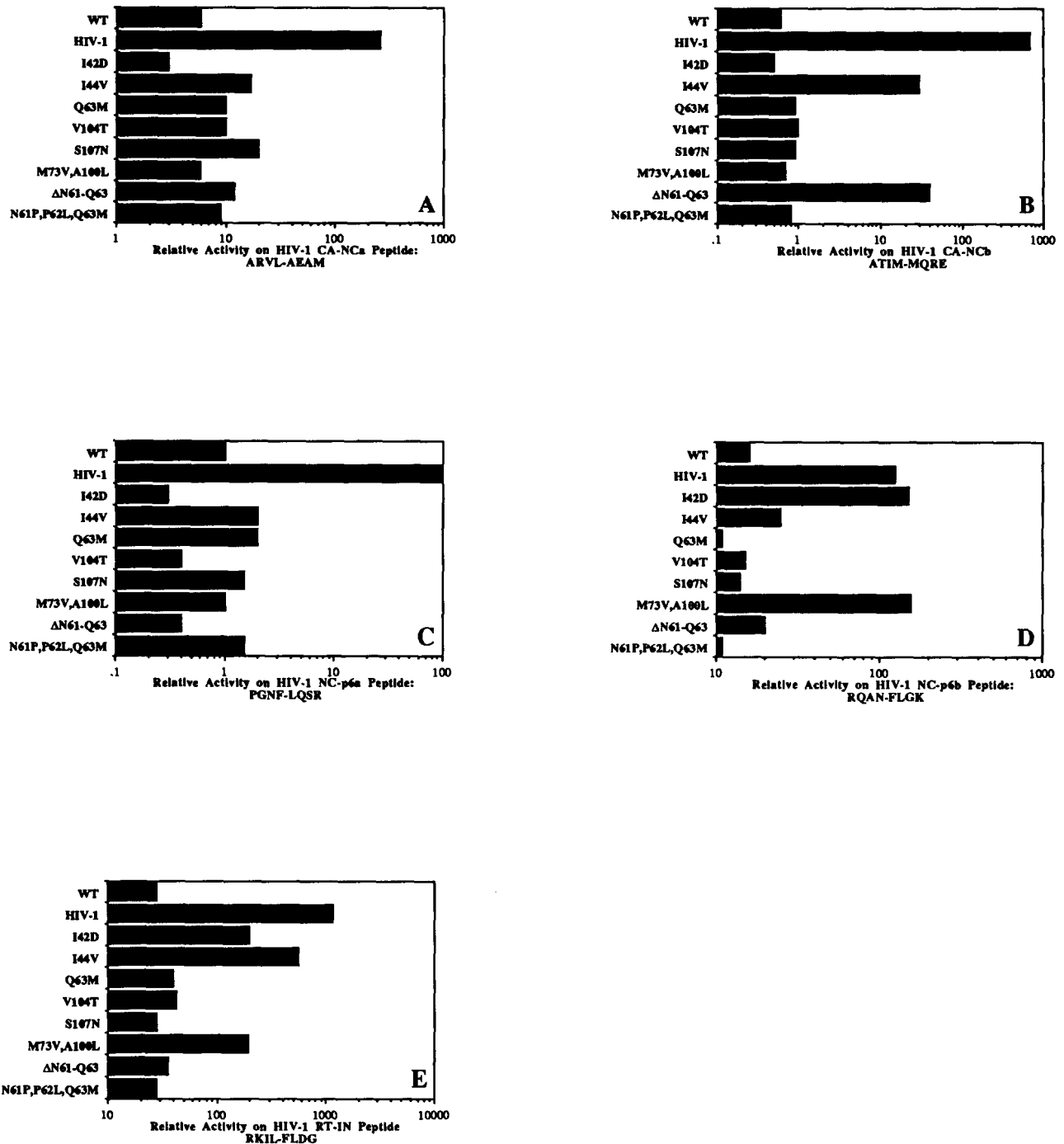
Mutational Analysis of Enzyme Subsites of RSV and HIV-1 PRs

FIG. 3. **Protease activities with HIV-1 peptide substrates.** Peptide substrates representing HIV-1 polyprotein cleavage sites CA-NCa (PARVL-AEAMRR) (*panel A*), CA-NCb (PATIM-MQRERR) (*panel B*), NC-p6a (PGNF-LQSRR) (*panel C*), NC-p6b (PRQAN-FLGKRR) (*panel D*), and RT-IN (PRKIL-FLDGRR) (*panel E*) were incubated with mutant PR (as indicated) and PR activity determined as described under "Experimental Procedures." The data are presented as relative activity to cleavage of the RSV NC-PR peptide substrate. The sequence notations are the same as in the legend to Fig. 2. *WT*, wild type.

interactions with both *Asp-42* and *Asp-42'*. This latter interpretation implies that in addition to the seven enzyme subsites already defined (9), an eighth subsite, S4', also functions in substrate selection by interacting with the P4' substrate position. This has been confirmed in separate experiments in which amino acid substitutions were placed into the P4' position of the NC-PR substrate and found to alter cleavage of these substrates by RSV PR (data not shown).

Of the various subsites of the PR, S2 appears to have the greatest influence on substrate selection. This can be explained by considering two features. First, S2 (and S2') is comparatively small, and the size of the amino acid that can bind easily is restricted. Second, as described by Gustchina *et al.* (24), P2 differs from other positions, including P2', in forming two relatively strong hydrogen bond interactions between its amide and carbonyl oxygen, the flap, and the conserved water molecule. This predominance of the S2 subsite can be seen in the HIV-1 peptide cleavage data presented in Fig. 3. Furthermore, differences between HIV-1 and HIV-2 PRs could be attributed to *Val-32* and *Ile-47*, respectively, in the S2 subsites (23), whereas

A

ILE44   VAL P2

B

VAL44   VAL P2
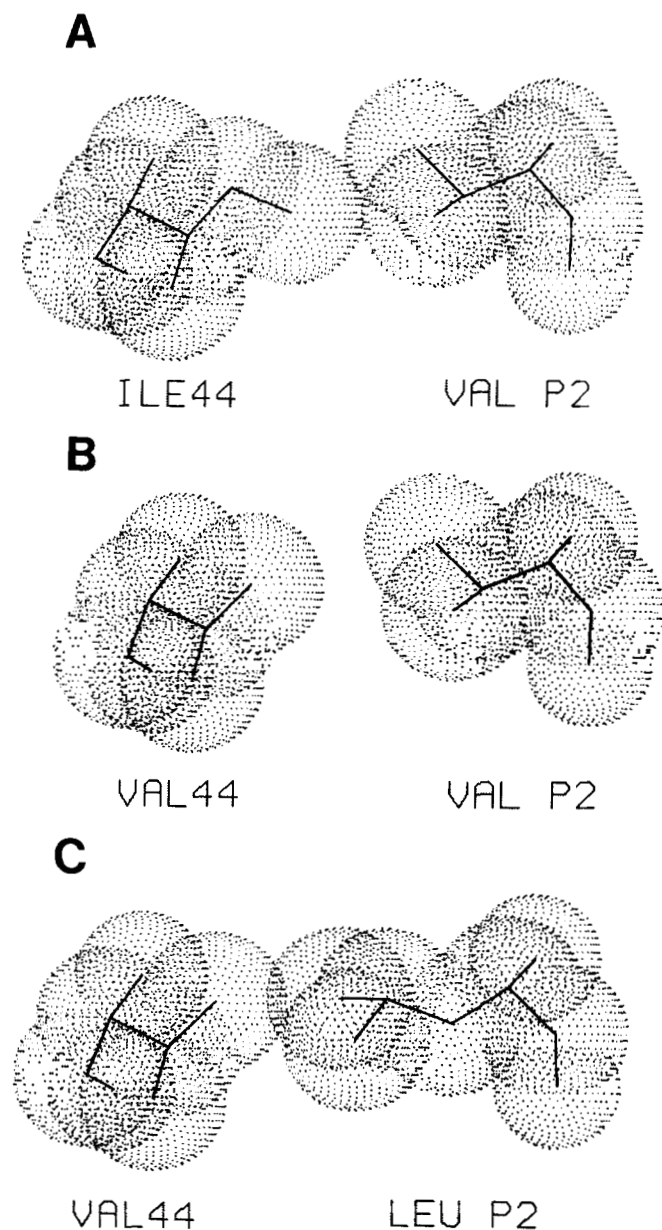
C

VAL44   LEU P2

FIG. 4. **Side chain interactions between the RSV residue 44 in the S2 subsite with the P2 position of the NC-PR substrate.** The van der Waal's interaction between RSV PR residue 44 and substrate residue P2 in the model structure is shown in a dot-surface representation. The side chains of the modeled residues have been rotated to give a good fit, the Ile-44 in the crystal structure was not moved. *Panel A*, wild type RSV protease Ile-44 from the crystal structure with valine in the P2 position of the NC-PR sequence. *Panel B*, RSV PR mutant containing the I44V substitution and valine in the P2 position of the NC-PR substrate. *Panel C*, RSV PR mutant with the I44V substitution and leucine in the P2 position of the NC-PR substrate. Note that these illustrations are from the model without energy minimization. In practice there would probably be some rearrangement of the protein and peptide structure to accommodate these residues. However, as this figure shows, any rearrangement to improve the contact for the *Val-44* with the valine at P2, *panel B*, would have to involve changes in the position of the main chain atoms and is likely to be more unfavorable than simply rotating the side chain.

differences between equine infectious anemia virus and HIV-1 PR were attributed to *D30T* and *I50V* changes also in the S2 subsite.

The small size of the retroviral PR, its broad substrate specificity, the relative ease at which it can be genetically manipulated and assayed with peptide substrates, and the availability of structural data make this an ideal model system for studying protein-protein interactions. Understanding the molecular basis of substrate selection for this system should provide information applicable to more complex biological problems including antibody interaction with an antigen and other cell surface components of the immune system.

REFERENCES

1. Skalka, A. M. (1988) *Cell* **56**, 911–913
2. Kohl, N., Emini, E., Schleif, W., and Davis, L. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 4185–4189
3. Crawford, S., and Goff, S. (1985) *J. Virol.* **53**, 899–907
4. Gottlinger, H., Sodroski, J., and Haseltine, W. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 5781–5785
5. Dickson, C., Eisenman, R., Fan, H., Hunter, E., and Teich, N. (1984) in *RNA Tumor Viruses* (Weiss, R., Teich, N., Varmus, H., and Coffin, J., eds) vol. 1, pp. 513–648, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
6. Cameron, C., Grinde, B., Jentoft, J., Leis, J., Weber, I. Copeland, T., and Wlodawer, A. (1992) *J. Biol. Chem.* **267**, 23735–23741
7. Darke, P., Nutt, R., Brady, S., Garsky, V., Ciccarone, T., Leu, C.-T., Lumma, P., Freidinger, R., Veber, D., and Signal, I. (1988) *Biochem. Biophys. Res. Commun.* **156**, 297–303
8. Kotler, M., Danho, W., Katz, R., Leis, J., and Skalka, A. M. (1989) *J. Biol. Chem.* **264**, 3428–3435
9. Cameron, C., Grinde, B., Jacques, P., Jentoft, J., Leis, J., Wlodawer, A., and Weber, I. (1993) *J. Biol. Chem.* **268**, 11711–11720
10. Grinde, B., Cameron, C., Leis, J., Weber, I., Wlodawer, A., Burstein, H., and Skalka, A. M. (1992) *J. Biol. Chem.* **267**, 9491–9498
11. Grinde, B., Cameron, C., Leis, J., Weber, I., Wlodawer, Burstein, H., Bizub, D., and Skalka, A. M. (1992) *J. Biol. Chem.* **267**, 9481–9490
12. Bizub, D., Weber, I., Cameron, C., Leis, J., and Skalka, A. M. (1991) *J. Biol. Chem.* **266**, 4951–4958
13. Alexander, F., Leis, J., Soltis, D., Crowl, R., Danho, W., Poonian, M., Pan, Y.-C., and Skalka, A. M. (1987) *J. Virol.* **61**, 534–542
14. Cha, S. (1975) *Biochem. Pharmacol.* **24**, 2177–2185
15. Jaskolski, M., Miller, M., Rao, M., Leis, J., and Wlodawer, A. (1990) *Biochemistry* **29**, 5889–5898
16. Miller, M., Schneider, J., Sathyanarayana, B., Toth, M., Marshall, G., Clawson, L., Selk, L., Kent, S., and Wlodawer, A. (1989) *Science* **246**, 1149–1152
17. Fitzgerald, P. M. D., McKeever, B. M., VanMiddlesworth, J. F., Springer, J. P., Heimbach, J. C., Leu, C.-T., Herbert, W. K., Dixon, R. A. F., and Darke, P. L. (1990) *J. Biol. Chem.* **265**, 14209–14219
18. Erickson, J., Neidhart, D., J., VanDrie, J., Kempt, D. J., Wang, X. C., Norbeck, D. W., Plattner, J. J., Rittenhouse, J. W., Turon, M., Wideburg, N., Kohlbrenner, W. E., Simmer, R., Helfrich, R., Poul, D. A., and Knigge, M. (1990) *Science* **249**, 527–533
19. Swain, A. L., Miller, M. M., Green, J., Rich, D. H., Kent, S. B. H., and Wlodawer, A. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 8805–8809
20. Jaskolski, M., Tomasselli, A., Sawyer, T., Staples, D., Heinrikson, R., Schneider, J., Kent, S., and Wlodawer, A. (1991) *Biochemistry* **30**, 1600–1609
21. Jones, A. (1978) *J. Appl. Crystallogr.* **11**, 268–272
22. Otto, M., Garber, S., Winslow, D., Reid, C., Aldrich, P., Jadhav, P., Patterson, C., Hodge, C., and Cheng, Y.-S. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 7543–7547
23. Tozser, J., Weber, I., Gustchina, A., Blaha, I., Copeland, T., Louis, J., and Oroszlan, S. (1992) *Biochemistry* **31**, 4793–4800
24. Gustchina, A., Sansom, C., Prevost, M., Richelle, J., Wodak, S., Wlodawer, A., and Weber, I. (1994) *Protein Eng.*, in press
25. Kent, S., Schneider, J., Selk, C., Delahunty, C., and Chen, Q. (1989) in *Current Communications in Molecular Biology: Viral Proteinases as Targets for Chemotherapy* (Krausslich, H.-G., Oroszlan, S., and Wimmer, E., eds) pp. 223–230, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY