Mutational sensitivity patterns define critical residues in the palm subdomain of the reverse transcriptase of human immunodeficiency virus type 1

Shih-Fong Chao⁺, Voon Loong Chan¹, Peter Juranka¹, Andrew H. Kaplan^{2,3,§}, Ronald Swanstrom^{2,4} and Clyde A. Hutchison III*

Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA, ¹Department of Microbiology, University of Toronto, Toronto M5S 1A8, Canada and ²Lineberger Comprehensive Cancer Center, ³Department of Medicine and ⁴Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Received November 10, 1994; Accepted January 27, 1995

ABSTRACT

We have analyzed 154 single amino acid replacement mutants within a 40 amino acid region (residues 164–203) of the reverse transcriptase (RT) from human immunodeficiency virus type 1 (HIV-1). This region consists of two antiparallel β -strands (strands 9 and 10) flanked by two α helices (E and F). The structure of this region of the 'palm' subdomain is conserved in a variety of DNA and RNA polymerases, indicating a critical role in enzyme structure and function. Functional assays were performed by screening RT activity of mutants expressed in *E.coli*. A functionally important region corresponding closely to β -strands 9 and 10 and the loop joining them was revealed by its mutational sensitivity. Structural analysis of mutants was performed by using Western blots to assay correct folding, which is required for processing to produce the mature p66 and p51 RT species. This analysis indicates that β -strand 10 is a structurally important region. Combined analysis of these two assays revealed diagnostic patterns of mutational sensitivity which identify key positions in the RT sequence at which a specific amino acid side chain is critical, either for structure or function, as well as residues which are external to the RT structure. This work illustrates the utility of large-scale mutagenesis in relating primary sequence to significant features of protein structure and function.

INTRODUCTION

We are using the saturation mutagenesis approach to study HIV-1 RT expressed in *Escherichia coli* (1), with the goal of gaining a more detailed understanding of the enzyme's structure and function. In this system the RT is first expressed as a large Pol

precursor and this precursor is processed to generate mature proteins, including the p66 and p51 chains of RT, indistinguishable from those found in virions (1–3) and infected cells (4). The processed RT is enzymatically active in crude bacterial extracts. This paper describes improvements in our strategy for saturation mutagenesis which we applied to a 40 amino acid region (amino acids 164–203). This region was originally selected for study because it contains the evolutionarily conserved sequence motif C (amino acids 178–191; 5) in its center. This motif is shared among several RNA and DNA polymerases, and contains a conserved Asp–Asp doublet (residues 185 and 186) found in many RNA-dependent polymerases including all known RTs. Previous mutational analyses of these conserved Asp residues suggest that they may play a catalytic role in polymerization by HIV-1 RT (6–10).

The availability of X-ray crystallographic structures for the HIV-1 RT (11,12) and several other polymerases has provided added interest in the region selected for our mutational analysis. The conserved Asp–Asp doublet is located at the junction of antiparallel β -strands 9 and 10 (Asp185 is in the loop and Asp186 at the end of β -strand 10). The region we have mutagenized lies within the 'palm' subdomain of the RT and consists of β -strands 9 and 10 and most of the two α helices (E and F) which flank them in the primary amino acid sequence. Most significantly, this structure consisting of two antiparallel β -strands flanked by two α helices constitutes the 'topologically conserved core' (13) of the structures of a diverse collection of DNA and RNA polymerases. In addition to its fundamental importance in understanding the structure and function of RT, this region of the HIV-1 RT is of interest as a target for HIV inhibitors. X-ray crystallographic data (11) indicates that this region of the HIV-1 RT interacts with the non-nucleoside inhibitor Nevirapine. Recent studies have shown that at least five residues in this region are involved in resistance to several drugs (14–18).

* To whom correspondence should be addressed

Present addresses: ⁺Division of Infectious Diseases, Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA and [§]37-121 CHS, UCLA School of Medicine, 10833 Leconte, Los Angeles, CA 90024, USA

Our goal in these studies was to examine a large collection of mutants, including single amino acid replacements at each position within the target region. To achieve this goal we used two simple phenotypic assays to analyze 154 different RT mutations. Our first assay was measurement of total RT activity in crude bacterial extracts induced to produce each mutant RT. Some mutations leading to reduced RT activity involve residues directly involved in catalysis by the enzyme. Other mutations lead to a loss in RT activity by disrupting enzyme structure. Western blot analysis of each mutant was used as a second assay to identify mutations which disrupt the RT structure. The assumption underlying this assay is that a correctly folded RT will be processed in *E. coli* by the linked viral protease to produce p66 and p51 in wild-type amounts (1). Mutations which reduce or eliminate correct processing are presumed to do so by interfering with correct folding, since they retain the wild-type protease gene and protease cleavage sites. Analysis of the results of these two assays allowed us to identify key amino acid residues and important regions within the palm subdomain of HIV-1 RT.

This work demonstrates how mutational studies can be combined with crystallographic data to gain insights into protein structure and function. Our data can be summarized in some simple rules relating the mutational sensitivity pattern of an individual residue to its functional and structural role. We expect that these methods for large-scale mutagenesis will also be useful in studying proteins for which detailed structural information is not available.

MATERIALS AND METHODS

Plasmid constructions

The HIV-1 pol gene used in these studies was originally derived from the cloned HXB2 provirus (19,36). Three silent mutations at amino acids 93, 94 and 205 were introduced into the HIV-1 RT expression plasmid pART1E66 (pE66) containing the HIV-1 pol gene (19), to create two new unique restriction enzyme cleavage sites, *Bam*HI and *Sac*I. These sites in the resulting plasmid (pE66M) define a cassette of 326 bp to which all mutations are confined (20). The 1.8 kb *Bam*HI-*Eco*RI segment of pE66M containing the cassette was subcloned into the *Bam*HI-*Eco*RI sites in the polylinker region of pUNC9(-) (21) to generate pCRT1.

Saturation mutagenesis of pCRT1

To saturate a selected region of the RT-coding sequence with mutations, eleven 36-base mutagenic oligos were designed which span the entire 326 bp mutagenic target. The mutagenic target defined by each oligo overlapped by six nucleotides with the adjacent oligo. An average of 1.5 random nucleotide mutations were incorporated per oligo during automated synthesis according to Hutchison *et al.* (22,23). Each of the 11 mutagenic oligos was used to generate a separate library with random point substitutions over its target region by using a uracil-containing, single-stranded DNA form of pCRT1 following the procedure of Kunkel *et al.* (24). The *in vitro* synthesis products were transformed into *E. coli* JM101 under the selection of kanamycin resistance, and colonies on the 2XYT kanamycin plate were pooled to give a mutant library for each mutagenic oligo in a pCRT1 genetic background.

Mutant phage library construction

Plasmid DNA from each pCRT1 mutant library was prepared, digested with *Bam*HI, *Sac*I and *Hae*III (*Hae*III was used to digest the competing large *Bam*HI-*Sac*I segment containing the vector; it does not cleave the 326 bp cassette), phenol extracted, and precipitated with ethanol. The resulting *Bam*HI-*Sac*I cassette from each pCRT1 mutant library was subcloned back into the expression vector, pE66M. To do this, double-stranded plasmid DNA of pE66M was prepared, digested with *Bam*HI and *Sac*I, and treated with alkaline phosphatase. After subcloning, a new library (in pE66M) was obtained by transforming *E. coli* JM101 and selecting for ampicillin resistance. A representative sample from each library was packaged as M13 phage particles by the M13K07 helper phage (25) to give a phagemid stock for each library which was used in subsequent experiments.

Genotype screening

After low multiplicity infection of *E. coli* JM101 with each phagemid library, individual clones were randomly picked, single-stranded DNA was prepared and sequenced (20,26). The resulting clones contained 0, 1, 2, 3 or more nucleotide substitutions, some of which cause amino acid substitutions. Clones with mutations that produce a single amino acid replacement were subjected to the following phenotypic assays. In general, we stopped collecting mutants for each library when there was at least one single missense mutation isolated for each codon spanned by the oligo.

Phenotype screening

RT mutant phagemid clones were used to infect *E. coli* JM101. Single colonies selected on glucose-minimal-ampicillin plates were used to inoculate two cultures, to confirm the mutant sequence (as above), and to determine its phenotype. Culture medium for phenotypic assays was M9 minimal medium plus 0.4% casamino acids, 0.001% thiamine and 50 µg/ml ampicillin. A 5 ml culture of each mutant was grown in a 50 ml tube at 37°C to an optical density of 0.3–0.4 at 600 nm, and cultures with equivalent amounts of cells were induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 37°C for 2 h. Then each induced culture was divided into two aliquots, one for Western blot as described by Loeb *et al.* (26), and the other for RT activity as described by Farmerie *et al.* (1) with modifications. In brief, RT activity was measured in 25 µl reactions with 1 or 0.1 µl of crude bacterial extract using poly(rC)-oligo(dG) as primer-template. RT activity assays were performed in duplicate in wells of a 96 well microtiter tray at 37°C for 10 min. Samples (5 µl) of RT reaction products were spotted on Whatman DE-81 filter, washed, and the incorporation of [³²P]dGMP into DNA was quantitated by the AMBIS Radioanalytic Imaging System. RT activity was measured under conditions where incorporation of labeled dGMP was linear with time for wild-type enzyme (in each assay, 1 and 0.1 µl of wild-type RT were analyzed in duplicate). RT activity of each mutant or wild-type clone was determined by subtracting the average RT activity of the background clones (bacteria containing the expression vector without the HIV-1 pol gene) from its own average RT activity. The fraction of wild-type RT activity for each mutant was calculated by dividing the RT activity of the mutant clone by the RT activity of the wild-type RT clone.

Mutant	RT assay/Western	Mutant	RT assay/Western	Mutant	RT assay/Western	Mutant	RT assay/Western	Mutant	RT assay/Western					
M164I	102 (+++)	+	K173N	95 (+++)	+	Y181H	15 (+)	+	L187S	0 (-)	-	G196V	39 (++)	+
164L	124 (+++)	+	173I	98 (+++)	+	Q182E	2 (-)	+/-	Y188F	43 (++)	+	196W	53 (+++)	+
164V	16 (+)	+	Q174E	133 (+++)	+	182P	1 (-)	+/-	188C	13 (+)	+/-	196R	28 (++)	+/-
164T	4 (+)	+/-	174K	95 (+++)	+	182R	0 (-)	+/-	V189I	20 (++)	+	Q197E	74 (+++)	+
T165S	146 (+++)	+	174P	5 (+)	+/-	182H	24 (++)	+	189L	33 (++)	+/-	197R	61 (+++)	+
K166R	52 (+++)	+/-	174R	89 (+++)	+	182K	0 (-)	+/-	189A	11 (+)	+/-	197L	52 (+++)	+
166T	41 (++)	+	174L	104 (+++)	+	Y183F	20 (++)	+	189G	0 (-)	-	197H	77 (+++)	+
I167L	40 (++)	+	N175K	1 (-)	+/-	183S	<1 (-)	+	G190A	75 (+++)	+	H198R	1 (-)	-
167M	50 (+++)	+	175T	5 (+)	+/-	183N	0 (-)	+	190R	2 (-)	+/-	198Y	2 (-)	-
167S	2 (-)	+/-	175S	8 (+)	+/-	183C	5 (+)	+	190E	4 (+)	+/-	198N	1 (-)	-
167T	4 (+)	-	P176A	61 (+++)	+/-	183D	0 (-)	+	S191A	2 (-)	-	198P	0 (-)	-
L168F	2 (-)	+/-	176L	35 (++)	+/-	183H	2 (-)	+	191P	0 (-)	-	R199T	37 (++)	+/-
168S	1 (-)	-	176R	55 (+++)	+	M184I	35 (++)	+	191F	0 (-)	-	199I	9 (+)	-
E169D	125 (+++)	+	D177E	135 (+++)	+	184L	25 (++)	+	191Y	0 (-)	-	199S	32 (++)	+/-
169V	19 (+)	+/-	177H	80 (+++)	+	184V	45 (++)	+/-	D192E	57 (+++)	+	T200A	111 (+++)	+
P170A	50 (+++)	+	177V	14 (+)	+/-	184K	0 (-)	+	192N	46 (++)	+	200S	65 (+++)	+
170S	45 (++)	+	177A	71 (+++)	+	D185E	1 (-)	+	192Y	6 (+)	+/-	200P	2 (-)	-
170H	20 (+)	+/-	I178V	15 (+)	-	185N	0 (-)	+	192V	4 (+)	+/-	200I	42 (++)	+/-
170L	25 (++)	+	178L	128 (+++)	+	185V	0 (-)	+/-	192G	21 (++)	+/-	200K	51 (+++)	+
170R	16 (+)	+/-	178T	48 (++)	+/-	185G	0 (-)	+	192A	14 (+)	+/-	200R	124 (+++)	+
F171Y	56 (+++)	+	178R	2 (-)	-	185Y	0 (-)	+	L193I	42 (++)	+/-	K201N	5 (+)	-
171C	2 (-)	-	V179D	145 (+++)	+	D186E	<1 (-)	+	193V	33 (++)	+/-	201I	53 (+++)	+
171I	4 (+)	-	179F	4 (+)	+/-	186N	0 (-)	+	193F	127 (+++)	+	I202V	108 (+++)	+
171L	7 (+)	+/-	I180M	20 (++)	+/-	186A	0 (-)	+	E194D	145 (+++)	+	202M	21 (++)	+
R172K	42 (++)	+	180L	54 (+++)	+	186H	0 (-)	+	194K	95 (+++)	+	202K	<1 (-)	-
172S	8 (+)	+	180F	9 (+)	+	186G	<1 (-)	+	194G	107 (+++)	+	202R	<1 (-)	-
172T	8 (+)	+	180T	3 (+)	+	186Y	0 (-)	+	194V	83 (+++)	+	E203D	89 (+++)	+
172G	3 (+)	+/-	180S	<1 (-)	-	L187V	14 (+)	+	I195L	115 (+++)	+	203V	58 (+++)	+
K173R	149 (+++)	+	180N	<1 (-)	-	187M	86 (+++)	+	195T	125 (+++)	+	203G	22 (++)	+/-
173T	86 (+++)	+	Y181F	13 (+)	+/-	187F	9 (+)	+/-	195K	96 (+++)	+	203A	55 (+++)	+
173Q	106 (+++)	+	181S	8 (+)	+/-	187W	0 (-)	+/-	195R	80 (+++)	+			

Figure 1. Summary of the RNA-dependent DNA polymerase (RT) assay and Western blot results. RT assay results are listed as percentage of the wild-type activity. Mutant phenotypes are classified as: negative (-) for <3% of the wild-type activity; low intermediate (+) for ≥3% of the wild-type activity; high intermediate (++) for ≥20% of the wild-type activity; and wild-type or positive (+++) for ≥50% of the wild-type activity. Some clones (24%) were assayed twice or more from lysates prepared independently to check the reliability of the data. RT Activities of duplicate experiments were within a factor of two, except for some lysates which had <3% of wild-type RT activity. Western blot results are analyzed for RT processing, and the phenotypes are classified as: wild-type or positive (+): efficient processing to produce stable products, as indicated by the presence of near wild-type levels of the p66 and p51 forms of the mature reverse transcriptase; negative (-): inefficient processing, indicated by the near absence of p51, a much reduced level of p66, with some high and low molecular weight products when compared to p66; or intermediate (+/-): this category contains the widest variety of phenotypes, generally indicated by a reduced level of the RT heterodimer. Some mutant clones (24%) were tested on two or more separate occasions from lysates prepared independently, and no discrepancies were observed.

RESULTS

Mutagenesis and assays

We improved our previously described strategy for saturation mutagenesis to avoid second site mutations located outside the region defined by the mutagenic oligo (23). The target region of the HIV-1 *pol* gene was first engineered to lie within a 326 bp restriction fragment cassette, in a phagemid vector. After mutagenesis using each randomly mutagenized oligo, a mutant library was prepared as a double-stranded DNA restriction fragment. This fragment was then cloned into a plasmid containing the remainder of the *pol* gene in a vector suitable for screening phenotypes. This confines all changes introduced during mutagenesis to the 326 bp, which can be sequenced from a single sequencing primer. A total of 154 different single amino acid substitutions spanning a 40 amino acid region were identified by sequencing, and listed in Figure 1.

We used two assays to determine the phenotype of each mutant (see Fig. 1 for a summary of results). The first assay measured RT activity at 37°C. Western blot analysis was used as a second phenotypic assay to examine processing and stability of mutant RT produced in *E.coli* by HIV-1 protease. In the wild-type RT clone (pE66M), the HIV-1 protease is expressed together with RT and integrase as a large Pol precursor. This precursor is processed to generate several *pol* gene products including the two forms of RT, p66 and p51, detectable by an HIV-1 RT specific antibody. Figure 2 shows the wild-type Western blot pattern, along with

representative examples of mutants with positive, negative and intermediate phenotypes.

Mutational effects on RT function: patterns of sensitivity

To analyze the effects of mutations on RT activity, mutant RTs were first classified into four phenotypic categories (-, +, ++, +++) based on their percentage of wild-type RT activity (see legend to Fig. 1).

From the data in Figure 1, patterns of sensitivity were analyzed at each residue within the target region. As expected, RT activity is generally more sensitive to single non-conservative amino acid substitutions than to single conservative amino acid substitutions. We find it useful to classify each sequence position into one of three different patterns of sensitivity to amino acid substitution:

Mutational insensitivity for catalytic activity. At certain residues RT retained wild-type activity with all single amino acid substitutions tested (both conservative and non-conservative). Mutationally insensitive positions are 173, 174 (except for a proline substitution), 194, 195 and 197.

Extreme mutational sensitivity for catalytic activity. In contrast, changes at several residues resulted in a negative phenotype for RT activity, for all single amino acid substitutions tested (both conservative and non-conservative). Extreme sensitivity was observed for glutamine at position 182 (except for a histidine substitution), aspartic acid at position 185, aspartic acid at position 186, serine at position 191 and histidine at position 198.

Figure 2. Examples of Western blot analysis of mutants. Western analyses demonstrating the three patterns of mutational sensitivity: a residue with all positive phenotypes, a residue with all negative phenotypes and a residue with mixed (positive, intermediate and negative) phenotypes. Crude extracts of bacteria expressing each mutant were prepared as described in Materials and Methods. Lanes: 1) molecular weight markers; 2) negative control (p120 which does not contain any HIV-1 sequence); 3) positive control (a wild-type HIV-1 RT clone); 4) I195L (HIV-1 wild-type RT with a I195 to L mutation) with a positive phenotype; 5) I195T [positive phenotype]; 6) I195K [positive phenotype]; 7) I195R [positive phenotype]; 8) S191A [negative phenotype]; 9) S191P [negative phenotype]; 10) S191F [negative phenotype]; 11) S191Y [negative phenotype]; 12) I178V [negative phenotype]; 13) I178L [positive phenotype]; 14) I178T [intermediate phenotype]; 15) I178R [negative phenotype]. p120 is the primary translation product of Pol precursor of pE66M which is used as the wild-type RT clone. p66 and p51, two processed products of p120, are the two forms of the HIV-1 RT.

These residues, except glutamine at position 182, are strictly conserved among HIV-1, HIV-2 and Visna RTs.

Mixed mutational sensitivity for catalytic activity. Most positions in the sequence are in this group which includes all positions which do not fall into either of the above categories. The amount of RT activity at these positions varies depending on the substitution.

Mutational effects on RT structure: patterns of sensitivity

To analyze the effects of mutations on protein folding, mutant RTs were first placed into three phenotypic categories based on their processing patterns in Western blots: wild-type or positive (+); negative phenotype(-); intermediate phenotype (+/-) (see legend to Fig. 1 for details). Analysis of Western blot data summarized in Figure 1 shows that 73% of the conservative amino acid substitutions had a positive processing phenotype, compared to only 47% of the single non-conservative amino acid substitutions. Conservative amino acid substitutions therefore appear less likely to disrupt correct protein folding than do non-conservative amino acid substitutions, as would be expected. We next classified each position within the target sequence according to its pattern of mutational sensitivity for protein structure, using a classification analogous to the one used for RT activity. Figure 2 shows Western analyses of mutants at sequence position which exhibit three different patterns of mutational sensitivity with regard to protein structure:

Mutational insensitivity for protein folding. Every change we observed at these positions (both conservative and non-conservative), resulted in a wild-type phenotype. This category

includes positions 173, 174 (except for a proline substitution), 183, 185 (except for a valine substitution), 186, 194, 195 and 197. This suggests that the amino acid residue at these positions is not critical for correct folding of RT in the *E.coli* environment.

Extreme mutational sensitivity for protein folding. Every change we observed at these positions (both conservative and non-conservative), resulted in a negative phenotype. Positions 191 (serine) and 198 (histidine) fall into this category. This suggests that the amino acid found at these positions in RT is critical for RT structure in the *E.coli* environment. Proteins from mutants at these positions probably do not fold correctly; therefore, the resulting conformation becomes inaccessible for correct processing but accessible to proteolytic degradation. As expected, these mutant RTs were also found to be negative for RT activity.

Mixed mutational sensitivity for protein folding. Most sequence positions fall into this category, which includes all positions not included in one of the categories above. Here, the extent of proper processing varies depending on the substitution.

Identification of a region important for catalytic activity

As an initial approach to analyze the information in Figure 1 we identified a mutationally sensitive region as previously described (26). To determine a mutationally sensitive region, we constructed a histogram of residue position versus mutant RT activity (not shown). For positions with mixed phenotype, the highest RT activity among the non-conservative substitutions at each position was plotted (26). As a result, we identified a mutationally sensitive region, amino acids 180–192. This mutationally sensitive region corresponds closely to conserved sequence motif C (amino acids 178–191), which consists of the two antiparallel β -strands 9 and 10 and the loop joining them in the protein structure (11,12; Figs 4 and 5).

A region important for structure

To look for regions of the sequence which are mutationally sensitive for RT structure, we constructed a histogram Western blot results for each sequence position (not shown). In this analysis the non-conservative substitution with the most efficient processing was plotted for positions with mixed phenotypes. In this way we identified a structurally important region from amino acid positions 187–192. This structurally important region, containing a structurally important amino acid at position 191, corresponds closely to β -strand 10 (11,12). β -strand 10 appears to interact with several β -strands and α helices, suggesting a physical basis for structural sensitivity of the mutations for this β -strand.

DISCUSSION

Saturation mutagenesis of the topologically conserved portion of the palm subdomain of HIV-1 RT

We employed saturation mutagenesis using randomly mutagenized oligonucleotides (22,23) to study 154 different HIV-1 RT mutants, including single amino acid replacements at each position for residues 164 through 203. Twenty different single amino acid replacements within our target region had been previously isolated using site-directed mutagenesis in studies by other workers (6–10,17,27). Eleven of these were also found in the collection presented here. Table 1 lists mutant RT activity

measurements reported previously and compares them with ours. In just three cases there appear to be significant differences (>2-fold) between our results and earlier studies. Our results are based on measurements on two independent clones of the same mutation in two of these cases. Differences in primer-template, preparation of RT extracts, and the bacterial expression system used may account for these quantitative differences. They do not affect any of our conclusions concerning mutationally sensitive regions or mutational sensitivity patterns of key residues.

Three-dimensional structure of the target region

A view of the X-ray crystallographic structure of the HIV-1 RT heterodimer complexed with DNA (12) is shown in Figure 5, with the mutagenic target region of the p66 and p51 subunits highlighted. An enlarged stereo view of the target region is also shown with key residues indicated, as identified by their mutational sensitivity patterns described below. These structural models show only the α -carbon backbone, as complete informa-

tion concerning orientation of side chains is not yet available for this structure (12). The target regions in the p51 and p66 subunits can be superimposed almost exactly (not shown). This suggests that mutations disrupting folding are likely to affect both subunits. Mutant phenotypes affecting catalytic activity are most likely expressed through the target region of the p66 subunit, because of its proximity to the DNA (12).

Diagnostic mutational sensitivity patterns

Mutational patterns of extreme sensitivity, sensitivity to non-conservative changes, or of insensitivity, for the two assays described above identify several different categories of key sequence positions. The mutational sensitivity patterns expected for residues of catalytic (C), functional (F) and structural (S) importance, and for external (E) residues, are indicated in Figure 3. Our rationale for this classification, and its application to the data is described below, and summarized in Figures 4 and 5.

Table 1. Mutants isolated by other workers within our target region

Mutant	RT(%)	Ref.	Comment	Mutant	RT(%)	Ref.	Comment
P170L	100	7		M184A	5–10	27	poly rA/U
	25	*	One clone		70–100	27	poly rA/dT
V179D	100	17		M184P	0	27	poly rA/U
	145	*	One clone		0	27	poly rA/dT
Y181D	1.3	8		D185H	<0.01	7	
Y181I	110	17		D185N	<0.01	10	
Y183F	30	6			<1	9	
	20	*	Two clones		0	*	One clone
Y183S	1.2	7		D185E	<0.01	10	
	<1	*	Two clones		<5	6	
M184L	5	8			1	*	Two clones
	8	10		D186N	<0.1	8	
	25	*	Two clones		<0.01	10	
M184V	WT	27	poly rA/U		<1	9	
	70–100	27	poly rA/dT		0	*	One clone
	45	*	Two clones	D186E	<0.01	10	
M184G	0	27	poly rA/U		<5	6	
	5–10	27	poly rA/dT		<1	*	Two clones
M184Y	20	6		Y188L	110	17	
M184S	5–10	27	poly rA/U	G190R	23	7	
	70–100	27	poly rA/dT		2	*	Two clones

Mutants designated in bold type were also studied in this paper.

Ref. 6: poly rC-oligo dG used as primer-template. RT assayed in bacterial extracts. Sequence coding for p66 plus two additional N-terminal amino acids is cloned in pUC.

Ref. 7: poly rA-oligo dT used as primer-template. RT assayed in bacterial extracts. Sequence coding for p66 plus three additional N-terminal amino acids was cloned in M13.

Ref. 8: Primer-template, RT purity in assay, and RT gene are the same as ref. 7.

Ref. 9: poly rC-oligo dG used as primer-template. RT is purified. p66 with six additional N-terminal histidine residues was expressed in *E.coli*.

Ref. 10: Primer-template, RT purity in assay, and RT gene appear to be the same as ref. 7.

Ref. 17: Conditions are the same as ref. 6.

Ref. 27: Use poly rA-oligo dT or poly rA-oligo U as primer-template. RT is partially purified. RT is processed by HIV-1 protease to generate mature heterodimer RT in *E.coli*.

Ref. *: This paper. poly rC-oligo dG used as primer-template. RT assayed in bacterial extracts is processed in *E.coli* as ref. 6 but with different construct.

		Folding	
		+	-
Catalytic Activity	+	E	(-)
	+/-	F	(-)
	-	C	S

Figure 3. Structural and functional significance of mutational sensitivity patterns. Positions which show mutational patterns of either extreme sensitivity (-), sensitivity to non-conservative changes (+/-), or insensitivity (+), for the two assays identify several different classes of functionally and structurally significant residues. 'E' indicates the pattern expected for external residues. 'C' and 'F' designate residues of catalytic and functional importance. 'S' indicates structurally important residues. Classes indicated by '(-)' are expected to be empty.

Catalytically and functionally important residues

If all mutations at a particular residue have a consistently wild-type phenotype for protein folding (HIV-1 protease processing gives near wild-type levels of p66 and p51 RT) but a

consistently negative phenotype for catalytic activity (<3% of the wild-type RT activity in the *in vitro* assay), this residue is likely to be involved directly in catalysis. We can identify two hydrophilic aspartic acid residues at positions 185 and 186 (near the junction of β -strands 9 and 10) as this type of catalytically important residue (designated 'C' in Figs 4 and 5). In these two positions, many types of side chain substitutions including non-polar, neutral polar, and charged side chains dramatically reduce enzyme activity. Therefore, the negatively charged side chains of these two aspartic acid residues at positions 185 and 186 of HIV-1 RT are required for catalytic activity. A catalytic role is also suggested for these residues given their extensive conservation (5) and proximity to the site of catalysis (12), however the analysis presented here suggests a strategy for assessing a role in catalysis based entirely on mutational sensitivity. It is interesting that in some proteins, mutation of catalytic residues can lead to significant structural changes, in apparent contrast with our findings for the HIV-1 RT (for review see 35).

We can identify other functionally important residues, defined by a mutational sensitivity pattern in which non-conservative amino acid substitutions have a severe effect on catalytic activity, but not on folding. These positions are indicated by 'F' in Figures 4 and 5. One such residue is arginine at position 172 (in α helix E). Non-conservative single amino acid substitutions at this position (except glycine) had wild-type processing phenotypes

Figure 4. Phenotypes of single missense mutations compared with the secondary structure for a 40 amino acid region. Mutants were isolated from four overlapping random substitution libraries designated as S, T, U and V. The region mutagenized in each library is indicated by a horizontal line. Residues are numbered from the N-terminus of the mature RT. Among the mutations isolated, approximately one third (52/154) were conservative amino acid changes, and about two thirds (102/154) were non-conservative changes. Underlined residues in the wild-type sequence are conserved in HIV type 1 (28), HIV type 2 (29), visna virus (30), bovine leukemia virus (31), Rous sarcoma virus (32) and murine leukemia virus (33). The conservative amino acid groups used in this analysis are as follows: (A, S, T, G, P); (V, L, I, M); (F, Y, W); (R, K, H); (D, E, N, Q); (C) (34). RT phenotypes are expressed as follows: upper-case = wild-type or positive; italicized upper-case = high intermediate; italicized lower-case = low intermediate; lower-case = negative. Those residues that are mutationally insensitive for catalytic activity are indicated by '0'. The Western phenotypes are expressed as follows: upper-case = wild-type or positive; italicized upper-case = intermediate; lower-case = negative. An 'X' on the line labeled 'Western sensitivity pattern' indicates that residue is mutationally insensitive for protein folding, while a '0' indicates a position of extreme mutational sensitivity for protein folding. The wild-type sequence is repeated with key residues designated beneath it as follows: F = functionally important; E = likely to be external; S = structurally important; C = catalytically important. The structural elements found in this region are indicated at the bottom of the figure (12).

Figure 5. Three dimensional structure of the mutagenized region of HIV-1 RT. The upper portion of the figure shows the α -carbon backbone of the RT heterodimer complexed with DNA (12). Positions of phosphorous atoms in DNA are indicated by crosses. Residues 160–210 in the two subunits are indicated by darker lines, and the mutagenized region of the p51 and p66 chains are indicated. The lower portion is a stereo view of the mutagenized region of p66, shown in the same orientation as the upper portion of the figure. Every tenth residue is numbered. The significance of key residues deduced from the mutagenesis studies is indicated using the same symbols as Figure 3. A line joining residues 191 and 198 indicates an interaction between the two corresponding side chains proposed on the basis of our results.

but low RT activity. Another functionally important amino acid is tyrosine at position 183 (at the end of β -strand 9, next to the loop). All non-conservative single amino acid substitutions at 183 had wild-type processing phenotypes but negative RT activity phenotypes (except Y183C which was 5% of wild-type activity). Methionine (M) at position 184 (in the loop joining β -strands 9 and 10) appears to be another functionally important residue, however we identified only one non-conservative amino acid substitution at this position (M184K). This mutant RT had a positive processing phenotype and was negative for RT activity. Interestingly, the four consecutive amino acids Y183, M184, D185 and D186 are identified by this analysis as functionally important residues. These residues constitute the most conserved motif characteristic of RTs (YMDD). This sequence includes the loop joining β -strands 9 and 10 (11,12). We have shown here that the RT structure is mutationally insensitive to many single amino acid substitutions within this motif. Preliminary analysis of the structure of RT suggests that the positions we have labeled as functionally important share the property of being partially buried (E. Arnold, personal communication).

Structurally important amino acids

If mutations at a particular residue have a consistently negative phenotype for protein folding, this position can be considered structurally important. Mutant RTs at these positions were poorly processed and gave much reduced levels of p66 and nearly undetectable p51 levels on Western blots. As would be expected, such mutants also are negative for RT activity. We identify serine at position 191 (in β -strand 10) and histidine at position 198 (in α helix F) as structurally important (indicated by 'S' in Figs 4 and 5). These residues are located close together in the three-dimensional structure of RT, and it is attractive to speculate that they interact to promote correct folding of β -strand 10 and α helix F.

The proximity of these α carbons (7.03 Å) suggests that H-bonding occurs between S191 and H198. Examination of unpublished X-ray structural data concerning location of these side chains shows that they do interact (E. Arnold, personal communication). This interaction is indicated by a line between positions 191 and 198 in the stereo view of Figure 5. A general feature of the residues we identified as structurally important is that they are buried and not exposed on the surface of the protein (E. Arnold, personal communication).

The glutamine at position 182 (in β -strand 9) appears to be a structurally important residue. At this position, all conservative and most non-conservative single amino acid substitutions (except histidine) resulted in intermediate phenotypes for protein folding and negative phenotypes for catalytic activity. However, histidine substitution at position 182 leads to a wild-type processing phenotype and high intermediate RT activity.

External amino acids

Mutants at certain positions have consistently wild-type phenotypes for both protein processing and RT activity. The fact that amino acid substitution at these positions does not interfere with RT folding or activity suggests that these residues are not in contact with other critical residues. This is the pattern expected for amino acids with side chains external to the structure. Our data show that the side chain of this kind of amino acid can be substituted with any other type of side chains except for proline. We identify five external amino acids in this 40 amino acid region; they are lysine at position 173, glutamine at position 174, glutamic acid at position 194, isoleucine at position 195 and glutamine at 197 (indicated by 'E' in Figs 4 and 5). Four of these are hydrophilic amino acids, and only one (isoleucine) is a hydrophobic amino acid. It should be noted that the residues identified as external by these mutagenesis studies do actually appear to be located on the surface of the protein, with side chains exposed (E. Arnold, personal communication).

Conclusions

The X-ray structure of HIV-1 RT is becoming available and is currently an area of active interest. When the mutagenesis data set we have generated is compared to the preliminary structure (including side chains), the fit is gratifying. The pattern of mutational sensitivity at a position has allowed us to identify positions that are likely to be important for catalysis, positions that affect protein folding and share the property of being buried in the protein structure, positions that do not play a major role in folding but affect activity and are partially buried, and residues that are insensitive to mutagenesis and are external in the structure. Mutations at residues identified as 'functional' by our analysis may, of course, exert their effects by structural perturbations which are more subtle than the block to folding caused by mutation of residues we classify as 'structural'. The ability to infer structural interactions critical for proteins folding from mutational analysis, such as the interaction between S191 and H198 in most cases will require at least a preliminary structural model. This illustrates the way that crystallographic and mutational data can complement each other to give a more detailed understanding of the roles of individual residues in protein structure and function. The methods for large-scale mutagenesis described here could also be applied to proteins for which structural data is unavailable. Rules relating the structural

and functional significance of a residue to its mutational sensitivity pattern (Fig. 3) may prove to be even more valuable in such cases.

ACKNOWLEDGEMENTS

We thank E. Arnold for discussing unpublished structural data, and for looking at positions of side chains of critical residues, to help interpret our data. We thank Susan Elmore for synthesis of oligonucleotides, and Y. D. Huang for technical assistance. This work was supported by Public Health Service Grants AI08998 and AI25321 to CAH and RS, and by NHRDP of Health and Welfare Canada to VLC. AHK was supported by a Scholar award from The American Foundation for AIDS Research, and by a Physician Scientist award from the NIAID (1956006143-AI).

REFERENCES

- Farmerie, W. G., Loeb, D.D., Casavant, N.C., Hutchison, C.A. III, Edgell, M.H. and Swanstrom, R. (1987) *Science* **236**, 305–308.
- Veronese, F. D., Copeland, T.D., DeVico, A.L., Rahman, R., Oroszlan, S., Gallo, R.C. and Sarngadharan, M.G. (1986) *Science* **231**, 1289–1291.
- Wondrak, E. M., Lower, J. and Kurth, R. (1986) *J. Gen. Virol.* **67**, 2791–2797.
- Lightfoote, M. M., Coligan, J.E., Folks, T.M., Fauci, A.S., Martin, M.A. and Venkatesan, S. (1986) *J. Virol.* **60**, 771–775.
- Delarue, M., Poch, O., Tordo, N., Moras, D. and Argos, P. (1990) *Protein Eng.* **3**, 461–467.
- Boyer, P. L., Ferris, A.L. and Hughes, S.H. (1992) *J. Virol.* **66**, 1031–1039.
- Larder, B.A., Purifoy, D.J.M., Powell, K.L. and Darby, G. (1987) *Nature (London)* **327**, 716–717.
- Larder, B.A., Kemp, S.D. and Purifoy, D.J.M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 4803–4807.
- Le Grice, S.F.J., Naas, T., Wohlgensinger, B. and Schatz, O. (1991) *EMBO J.* **10**, 3905–3911.
- Lowe, D.M., Parmar, V., Kemp, S.D. and Larder, B.A. (1991) *FEBS Lett.* **282**, 231–234.
- Kohlstaedt, L. A., Wang, J., Friedman, J.M., Rice, P.A. and Steitz, T.A. (1992) *Science* **256**, 1783–1790.
- Jacobo-Molina, A., Ding, J., Nanni, R.G., Clark, A.D.Jr., Lu, X., Tantillo, C., Williams, R.L., Kamer, G., Ferris, A.L., Clark, P., Hizi, A., Hughes, S.H. and Arnold, E. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 6320–6324.
- Sawaya, M.R., Pelletier, H., Kumar, A., Wilson, S.H. and Kraut, J. (1994) *Science* **264**, 1930–1935.
- De Vreese, k., Debyser, Z., Vandamme, A.-M., Pauwels, R., Desmyter, J., De Clercq, E. and Anne, J. (1992) *Virology* **188**, 900–904.
- Nunberg, J.H., Schleif, W.A., Boots, E.J., O'Brien, J.A., Quintero, J.C., Hoffman, J.M., Emini, E.A. and Goldman, M.E. (1991) *J. Virol.* **65**, 4887–4892.
- Shih, C.-K., Rose, J.M., Hansen, G.L., Wu, J.C., Bacolla, A. and Griffin, J.A. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9878–9882.
- Boyer, P.L., Currens, M.J., McMahon, J.B., Boyd, M.R. and Hughes, S.H. (1993) *J. Virol.* **67**, 2412–2420.
- Gu, Z., Gao, Q., Li, X., Parniak, M.A. and Wainberg, M.A. (1992) *J. Virology* **66**, 7128–7135.
- Erickson-Viitanen, S., Manfredi, J., Viitanen, P., Tribe, D.E., Tritch, R., Hutchison, C.A. III, Loeb, D.D. and Swanstrom, R. (1989) *AIDS Res. Hum. Retroviruses* **5**, 577–591.
- Chao, S.-F. (1992) Ph.D. Thesis, The University of North Carolina, Chapel Hill.
- Davies, C. J. and Hutchison, C.A. III (1991) *Nucleic Acids Res.* **19**, 5731–5738.
- Hutchison, C. A. III, Nordeen, S.K., Vogt, K. and Edgell, M.H. (1986) *Proc. Natl. Acad. Sci. USA* **83**: 710–714.
- Hutchison, C. A. III, Swanstrom, R. and Loeb, D.D. (1991) *Methods Enzymol.* **202**, 356–390.
- Kunkel, T. A., Roberts, J.D. and Zakour, R.A. (1987) *Methods Enzymol.* **154**, 367–382.
- Vieira, J. and Messing, J. (1987) *Methods Enzymol.* **153**, 3–11.
- Loeb, D. D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S.E. and Hutchison, C.A. III (1989) *Nature (London)* **340**, 397–400.
- Wakefield, J.K., Jablonski, S.A. and Morrow, C.D. (1992) *J. Virol.* **66**, 6806–6812.
- Ratner, L., Fisher, A., Jagodzinski, L.L., Mitsuya, H., Liou, R.S., Gallo, R.C. and Wong-Staal, F. (1987) *AIDS Res. Hum. Retroviruses* **3**, 57–69.
- Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L. and Alizon, M. (1987) *Nature (London)* **326**, 662–669.
- Sonigo, P., Alizon, M., Staskus, K., Klatzmann, D., Cole, S., Danos, O., Retzel, E., Tiollais, P., Hease, A. and Wain-Hobson, S. (1985) *Cell* **42**, 369–382.
- Sagata, N., Yasunaga, T., Tsuzuku-Kawamura, J., Ohishi, K., Ogawa, Y. and Ikawa, Y. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 677–681.
- Schwartz, D. E., Tizard, R. and Gilbert, W. (1983) *Cell* **32**, 853–869.
- Shinnick, T. M., Lerner, R.A. and Sutcliffe, J.G. (1981) *Nature (London)* **293**, 543–548.
- Dayhoff, M. O., Schwartz, R.M. and Orcutt, B.C. (1978) p.345–352. In M. O. Dayhoff (ed.), *Atlas of Protein Sequence and Structure*, vol. 5. National Biomedical Research Foundation, Washington, DC.
- Bone, R. and Agard, D.A. (1991) *Methods Enzymol.* **202**, 643–671.
- Ratner, L., Haseltine, W., Patarca, R., Livak, K.J., Starcich, B., Josephs, S.F., Doran, E.R., Rafalski, J.A., Whitehorn, E.A., Baumeister, K., Ivanoff, L., Petteway, S.F. Jr., Pearson, M.L., Lautenberger, J.A., Papas, T.S., Ghayeb, J., Chang, N.T., Gallo, R.C., and Wong-Staal, F. (1985) *Nature (London)* **313**, 277–284.