

ABSTRACT 629 : EXPANDED RANGE OF HIV-1 REVERSE TRANSCRIPTASE MUTATIONS DETECTED THROUGH LONG RANGE SEQUENCING

R. Kagan¹, M. Winters², T. Merigan² and P. Heseltine¹

¹Quest Diagnostics Nichols Institute, San Juan Capistrano CA 92690, ²Center for AIDS Research, Stanford University, Stanford, CA 94305

11th Conference on Retroviruses and Opportunistic Infections, 8-11 February 2004, San Francisco, CA

Abstract

Background: Most HIV-1 resistance-associated reverse transcriptase mutations are found within the N-terminal 240 residues of the RT enzyme. To search for additional mutational signatures outside the conventional range of many commercial genotyping assays, we applied a recently published statistical approach to a large clinical database of sequences spanning RT residues 1-400.

Methods: HIV-1 RT mutations spanning positions 1-400 from 41,122 deidentified clinical samples sequenced at Quest Diagnostics between 1/2002 and 6/2003 and 16,449 sequences from 7/2003 to 12/2003 were examined. Viruses with at least one resistance mutation were compared to those with no associated resistance using Chi-square statistics with the Benjamini-Hochberg correction for multiple comparisons. The binomial correlation coefficient Phi was calculated for all pairwise combinations of significant mutations. Clusters were identified with the Neighbor-Joining and the Fitch-Margoliash methods.

Results: The larger dataset contained 26,655 (64.8%) viral sequences with at least one predicted ARV resistance (PR or RT) and 14,467 (35.2%) sequences with no predicted resistances. Mutations at 57 RT positions (present in >1% of the ARV-resistant group) were found 2 times more often than in the latter group ($p < .01$ corrected for 400 tests). 37/57 positions were known primary or accessory resistance mutations, including nine recently-described NRTI associated mutations (39,43,203,208,218,221,223,228). Twenty novel mutations (28, 31, 40, 68, 90, 139, 172, 189, 224, 227, 237, 238, 242, 284, 348, 350, 355, 359, 370, 371) were also more frequent in the ARV-resistant group. Binomial correlation coefficients for 1,596 pairs identified three principal NRTI-related clusters and one NNRTI-related cluster. Residues 68,203,208,218,223 and 228 were assigned to the NRTI clusters and residues 221, 227 and 238 grouped with NNRTI clusters. Other significant correlations included residues 359+371 ($\phi = .372$), 181+221 ($\phi = .323$) and 106+227 ($\phi = .268$).

Conclusions: We have identified several novel RT mutations and associated them with known clusters of resistance related mutations in a large clinical data set. Several mutations did not group with known clusters, including five mutations in the RT connector domain. Statistical analysis of extended-length clinical RT sequences is a useful technique to uncover genotypic correlates of resistance. Phenotypic and clinical studies are needed to further elucidate the impact of these mutations on clinical response.

Introduction

Most resistance associated mutations to antiretroviral drugs in the HIV-1 reverse transcriptase (RT) are located within the N-terminal 240 residues of the RT enzyme, which includes the fingers and the palm domain of the p66 subunit (1, 2).

Commercial sequencing assays for the detection of drug resistance mutations vary in the length of RT cDNA sequenced, from codons 40-240 (Bayer TruGene™) to 1-335 (Abbott Viroseq™) or 1-400 (Tibotec-Virco *VirtualPhenotype*™). A recent study of the effect of RT sequence length on resistance predictions found that truncated RT sequence lengths to codon 240 affected 0.6% of *VirtualPhenotype*™ predictions, primarily due to omission of codons 318 and 333 (3).

Statistical methods may be used to identify mutational clusters correlated to nucleoside RT inhibitor (NRTI) resistance (4). In this work, we adapted such an approach to a large clinical database of RT sequences spanning codons 1-400 to identify additional mutations correlated with resistant genotypes both within and outside the conventional range of resistance assays.

Methods

A database of HIV-1 RT sequences spanning codons 1-400 from 41,122 deidentified clinical samples sequenced at Quest Diagnostics between 1/2002 and 6/2003 was partitioned into samples with predicted resistance to at least one ARV (NRTI or PI) and samples with no predicted resistances. We utilized the Chi-square test to identify positions present in > 1% of the ARV-resistant group that were present at a statistically higher frequency than in the ARV-sensitive group. To correct for multiple comparisons, we utilized the Benjamini-Hochberg correction (5) for 400 tests and a false discovery rate (FDR) set to 0.01.

We then calculated the binomial correlation coefficient phi, for the co-occurrence of two mutations in the same variant, using all pairwise combinations of these RT positions. We defined a distance metric of 1 - phi and constructed a pairwise distance matrix from the data set. We utilized the Fitch program (Dr. J. Felsenstein, University of Washington, Dept. of Genome Sciences) to run the Fitch-Margoliash least-squares method (6) with global rearrangements and three randomizations of the input distance to construct dendrograms for the identification of clusters of RT mutations. Dendrograms were output with TreeExplorer 2.12 (Dr. K. Tamura, Tokyo Metropolitan University, Tokyo, Japan). To assess the reproducibility of the clustering, we repeated the entire mutation identification and clustering procedure on a second independent data set of 16,449 sequences collected between 7/2003 and 12/2003.

Table 2: RT resistance associated mutational clusters¹

Mutation clusters	Known primary or accessory resistance mutations	Novel mutations
NRTI Cluster I	41, 44, 118, 210, 215	39, 43, 203, 223
NRTI Cluster II	67, 69, 70, 219	218, 228
NRTI Cluster III	62, 65, 68, 75, 77, 116, 151	
NNRTI Cluster I	101, 108, 181, 190	221
NNRTI Cluster II	106, 179, 188	227
NNRTI Cluster III	100, 103, 225	238
Uncharacterized cluster		359, 371
Unclustered resistance mutations	74, 184, 230, 318	

Unclustered novel mutations 28, 31, 40, 90, 139, 172, 189, 237, 242, 284, 348, 370

¹ Clusters were derived from the binomial correlation trees (1 - phi) shown in Fig. 1 and 2

Table 1: RT mutations that appear more frequently in ARV-resistant variants¹

RT position ²	Data Set I ³			Data Set II ³		
	Frequency	OR (95% CI)	Rank ⁴	Frequency	OR (95% CI)	Rank ⁴
28	2.6%	5.6 (5.3-6.0)	42	2.4%	4.7 (4.1-5.2)	44
31	1.3%	5.3 (4.9-5.9)	50	1.6%	4.2 (3.6-4.9)	51
39	13.7%	2.6 (2.5-2.7)	40	13.7%	2.5 (2.4-2.7)	38
40	1.4%	2.4 (2.2-2.6)	77	1.4%	1.7 (1.4-2.0)	113
41	30.8%	187 (166-189)	3	30.0%	183 (166-189)	3
43	11.0%	22.1 (21.6-22.7)	17	11.0%	15.8 (16.1-16.5)	19
44	8.2%	18.5 (17.9-19.0)	23	7.8%	27.5 (26.2-28.8)	15
62	3.3%	12.5 (11.9-13.2)	34	3.4%	15.2 (14.0-16.5)	29
65	3.0%	71.6 (69.0-74.2)	19	5.0%	87.4 (83.6-91.3)	12
67	31.7%	102 (101-103)	7	29.2%	159 (156-161)	5
68	6.9%	2.4 (2.3-2.5)	48	7.8%	2.8 (2.6-3.0)	40
69	16.3%	7.8 (7.4-7.7)	24	15.8%	6.4 (6.1-6.7)	24
70	22.6%	218 (216-220)	4	21.1%	136 (133-138)	7
74	13.0%	208 (206-211)	8	12.2%	144 (140-147)	8
75	5.9%	32.7 (31.7-33.7)	20	5.4%	23.8 (22.4-25.3)	20
77	1.2%	34.5 (32.2-37.0)	33	0.9%	13.1 (11.1-15.4)	39
90	5.3%	2.2 (2.1-2.3)	51	5.8%	2.4 (2.2-2.6)	46
100	4.4%	90.3 (87.9-92.8)	12	4.0%	47.0 (44.3-49.8)	16
103	4.8%	7.8 (7.4-7.9)	29	7.8 (7.4-7.9)	29	28
108	40.8%	23.2 (22.9-23.5)	10	40.8%	21.2 (20.7-21.6)	10
109	4.2%	2.3 (2.1-2.4)	56	4.6%	2.3 (2.3-2.8)	49
108	6.4%	8.5 (8.1-8.8)	31	6.4%	8.5 (8.0-8.1)	30
115	1.9%	46.6 (44.3-49.0)	26	2.3%	41.3 (38.2-44.7)	23
116	1.7%	48.6 (46.1-51.3)	27	1.0%	18.0 (15.6-20.8)	36
118	15.1%	5.8 (5.6-5.9)	30	15.4%	5.8 (5.6-6.1)	26
139	4.8%	117 (114-119)	11	5.2%	123 (119-128)	11
151	2.2%	64.2 (61.5-66.9)	22	1.5%	17.5 (15.5-19.8)	34
172	1.4%	3.7 (3.3-4.0)	58	1.7%	3.3 (2.9-3.9)	56
179	12.5%	2.2 (2.2-2.3)	43	13.5%	2.2 (2.1-2.3)	41
181	16.0%	289 (286-292)	5	15.3%	216 (212-220)	6
184	55.1%	1,329 (1,324-1,333)	1	54.4%	640 (636-645)	1
188	4.8%	117 (114-119)	11	5.2%	123 (119-128)	11
189	1.8%	1.9 (1.8-2.1)	84	2.0%	2.9 (2.5-3.3)	57
190	15.0%	77.6 (76.4-78.8)	9	14.6%	115 (112-117)	9
203	8.9%	6.5 (6.3-6.8)	32	8.0%	6.8 (6.4-7.2)	31
208	6.6%	23.9 (23.1-24.6)	21	6.2%	20.8 (19.7-22.0)	21
210	21.9%	15.4 (15.1-15.7)	14	20.5%	18.8 (18.2-19.4)	13
215	47.9%	88.8 (88.1-89.6)	6	45.3%	103 (102-105)	4
218	5.2%	56.1 (56.5-59.8)	15	4.9%	89.4 (86.2-92.8)	14
219	25.3%	457 (454-469)	2	23.9%	281 (277-284)	2
221	5.2%	44.0 (42.7-45.4)	18	5.1%	21.0 (19.8-22.4)	22
223	3.2%	28.9 (27.7-30.2)	25	3.0%	19.5 (17.9-21.1)	27
224	1.2%	2.2 (1.9-2.4)	90	1.2%	2.5 (2.1-3.0)	74
225	3.9%	20.8 (20.0-21.7)	28	4.6%	12.9 (12.0-13.8)	28
227	1.2%	15.0 (13.9-16.3)	38	1.3%	8.4 (7.3-9.7)	42
228	14.1%	23.4 (22.9-23.9)	13	13.5%	13.6 (12.9-14.2)	18
230	0.8%	3.2 (2.8-3.8)	71	1.2%	2.9 (2.4-3.4)	70
237	1.6%	5.9 (5.4-6.4)	47	1.3%	3.4 (2.9-4.0)	60
238	4.2%	6.3 (6.0-6.6)	37	4.0%	5.0 (4.5-5.4)	37
242	1.0%	15.9 (14.5-17.3)	41	1.1%	6.2 (5.2-7.4)	48
284	5.8%	2.9 (2.8-3.0)	45	5.4%	2.5 (2.3-2.7)	47
318	1.5%	11.3 (10.5-12.3)	39	1.6%	6.8 (5.9-7.8)	43
348	13.6%	20.1 (19.7-20.6)	16	13.0%	14.5 (13.9-15.1)	17
350	1.2%	2.6 (2.4-2.9)	72	1.2%	2.5 (2.1-3.0)	75
355	1.5%	2.0 (1.8-2.2)	89	1.5%	2.1 (1.8-2.4)	82
359	17.8%	2.8 (2.7-2.9)	36	17.7%	3.3 (3.2-3.5)	33
371	15.5%	3.9 (3.8-4.0)	35	15.7%	4.2 (4.0-4.4)	32

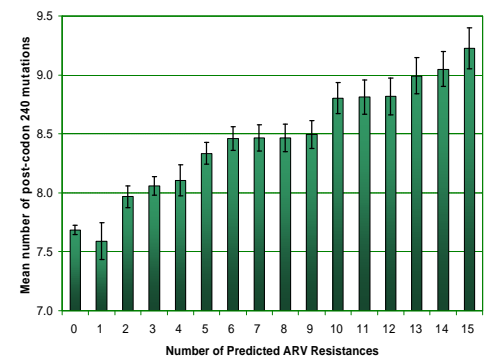
¹ The frequency of mutations in RT codons 1-400 was tabulated for ARV-resistant (at least one genotypically predicted resistance to any ARV) and ARV-sensitive (no predicted ARV resistances) viruses. 57 positions with a > 1% incidence of mutations were found more frequently in the first group, than in the second group (Odds ratio (OR) > 2 and Chi-square value significant at the False Discovery Rate (FDR) level of 0.01 (5)).

² Data Set 1: 41,122 sequences (64.6% with at least one ARV resistance) tested between 1/2002 and 6/2003

³ Data Set 2: 16,449 sequences (57.0% with at least one ARV resistance) tested between 7/2002 and 12/2003

⁴ Known primary or secondary resistance associated positions are colored in red.
⁵ Rank of Chi-square value for the mutated positions (1 = highest significance). The Benjamini-Hochberg FDR was calculated as 0.01 / (400 - rank).

Fig. 3 - Mean number of post-codon 240 RT mutations as a function of the number of predicted ARV resistances¹



¹ Mean number of mutations between RT codons 240 and 400 in 54,227 sequences with a minimum length of 1,410 nt, were tabulated, according to the number of ARV resistance predictions. Codon 318, used to predict DLV resistance, and mixed wt/mutant codons were not tabulated. The error bars represent the 95% CI of the mean.

Fig. 1 - A dendrogram of mutational clusters in data set I

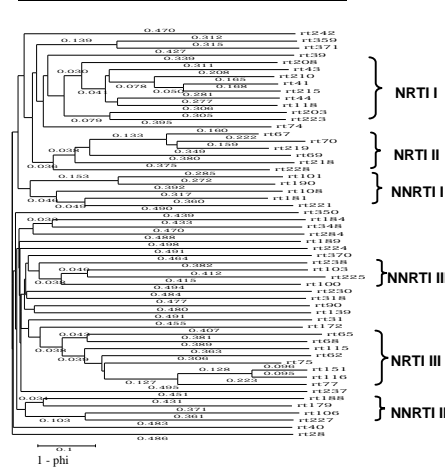
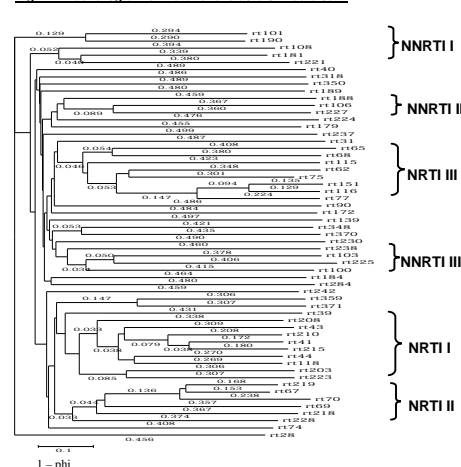


Fig. 2 - A dendrogram of mutational clusters in data set II



Results

There were 26,655 sequences (64.6%) in the 18-month dataset resistant to ≥ 1 ARV. Fifty-three positions appeared more frequently in this dataset (FDR < 0.01). Analysis of the second dataset (57% ARV-resistant variants) uncovered another four significant positions. Twenty three known primary or secondary NRTI resistance mutations and nine NNRTI resistance mutations were detected (Table 1). An additional eight mutations (39, 43, 203, 208, 218, 221, 223, 228) identified as significant were previously found to be associated with NRTI therapy by a similar method (4).

To determine the associations between the newly described mutations and the known resistance mutations, we calculated the binomial correlation coefficients for the 1,596 possible pairs of mutations. We found 252 pairs present more frequently than would be expected from the frequencies of the individual substitutions (Chi square test and FDR = 0.05). The median value of phi was 0.165 (range: 0.093 to 0.809). The second data set yielded similar rankings for 262 significant pairs with a median phi correlation of 0.159 (range: 0.094 to 0.736).

The co-mutation correlations (1 - phi) were input into the Fitch program to create dendrograms in which the branch lengths are proportional to the degree of co-mutation. The topology of the dendrograms obtained from both data sets was very similar (Fig. 1-2). The dendrograms show three principal NRTI clusters (Table 2) including two known thymidine analog mutation clusters (NRTI I and II) and the Q151M multidrug resistance cluster (cluster III). Four novel mutations clustered with NRTI cluster I (39, 43, 203 and 223; Table 2 and Fig. 1-2). Two additional mutations (218 and 228) clustered with NRTI cluster II (Table 2 and Fig. 1-2).

Principal Component Analysis by other investigators of 27 RT mutations associated with NRTI therapy (4) detected three clusters with comparable associations of the novel mutations. The amino acid substitution patterns for the novel mutations in both clusters were strongly biased towards specific residues and only codon 39 showed a higher degree of polymorphism (Table 3). A decision tree analysis of the TORO clinical trial data set placed the T39A/E mutation in the dL resistance decision tree further supporting a possible role for this mutation in NRTI resistance (7).

The non nucleoside RT inhibitor (NNRTI) mutations also formed three clusters (Table 2 and Fig. 1-2). Three novel mutations could be grouped with these clusters (Table 2). Mutations at codon 221 were previously found to be associated with NNRTI treatment (4). The F227L mutation in NNRTI cluster II forms part of the NNRTI binding pocket and can arise in cell cultures treated with the experimental NNRTI UC-781 (8). F227L was also predictive of DLV resistance in the TORO study decision tree analysis (7). The substitution patterns for positions 221 and 227 were strongly biased (H221Y and F227L) but codon 238 was polymorphic (Table 3).

The number of amino acid substitutions at RT positions C-terminal to codon 240 increased as a function of the number of predicted resistances to ARVs (Fig. 3), suggesting that mutations in this region may be associated with ARV resistance. However, the eight positions C-terminal to codon 240 located in the thumb or the connection domain did not cluster with the known resistance mutations (Tables 2, 3). Substitutions at positions 359 and 371 were correlated ($\phi = 0.373$; Fig. 1, 2) with each other, but not with other positions. G359 interacts with the primer strand near the RNase H domain and the G359S mutation has been observed to impair RNase H activity, resulting in lower viral titers in culture (9, 10). The N348I mutation in the connector domain is present in more than 13% of the resistant samples 14x to 20x more frequently than in the ARV-sensitive data set (Table 1). It's function, however, is not known.

Table 3: Amino acid substitution patterns and treatment associations of novel RT mutations

Cluster ¹	Position	Most common substitution(s) ²	RTI treatment association	RT protein domain	
NRTI I	39	T39A (78.4%); T39E/K (10.6%); E203K/D (89%)	NRTI ²	Fingers	
	43	K43E/Q/N (93.2%)	NRTI ²	Fingers	
	203	E203K/D (89%)	NRTI ²	Palm	
	223	K223Q/E (86.8%); K223T (10.6%)	NRTI ² , NNRTI ³	Palm	
NRTI II	218	D218E (99.8%)	NRTI ²	Palm	
	228	L228H/R (93.1%)	NRTI ²	Palm	
NNRTI I	221	H221Y (97.4%)	NRTI ² , NNRTI ³	Palm	
	NNRTI II	227	F227L (88.6%)	NNRTI ³	Palm
NNRTI III		238	K238T (70.9%); K238R (15.2%); K238N (10.5%)	Unknown	Palm
	Post-codon 240	242	Q242H (64.6%); Q242L (16.3%); Q242K (14.2%)	Unknown	Palm
		284	R284K (99.8%)	Unknown	Thumb
348		N348I (94.2%)	Unknown	Connection	
359	359	G359S (94.5%)	Unknown	Connection ⁵	
	370	E370D (52.8%); E370G (27.3%)	Unknown	Connection	
	371	A371V (92.7%)	Unknown	Connection	

¹ Clusters described in Table 2

² Quest Diagnostics database, 18-month data set 1 of resistant samples

³ Gonzales et al 2003

⁴ F227 is located in the NNRTI binding site; F227L has been observed in patients receiving the NNRTI UC-781 (8)

⁵ G359 interacts with the primer strand near the RNase H domain; the G359S mutation may impair RNase H activity and viruses carrying this mutation have a lower viral titer (9, 10).

Summary

► We have identified 20 novel RT mutations with a higher incidence in ARV-resistant viruses in a large clinical data set, and corroborated an association with resistance for an additional nine novel mutations in agreement with a previous study associating these mutations with NRTI therapy.

► Extended range sequencing identified eight mutations outside of the sequenced region of some commercial sequencing assays, that were associated with ARV resistant strains. A subset of these mutations in the RT connector domain may have a role in the binding and the positioning of the nucleic acid substrate or may affect RNase H activity.

► We have identified six clusters of known RT inhibitor resistance mutations that enabled preliminary classification of nine novel mutations as either putative NRTI or NNRTI-associated mutations. Mutational clusters may represent distinct evolutionary pathways of the RT in response to the selective pressure of RT inhibitors. Further statistical analysis and clustering of mutations in clinically characterized data sets may uncover additional relationships between these mutations and known patterns of resistance. Phenotypic and clinical studies are needed to further elucidate the impact of these mutations on clinical response.

REFERENCES

- Shaller RW, Kantor R, Gonzalez J. The genetic basis of HIV-1 reverse transcriptase and protease inhibitors. AIDS Rev. 2000; 2:211-228
- Shaller RW. Genotypic testing for Human Immunodeficiency Virus Type 1 drug resistance. Clin Microbiol Rev. 2002; 15(2):247-277
- Licozzi P, Van Houtte M, Veldman J and Bachevalier L. Impact of shorter sequence length on HIV-1 drug resistance analysis outcome and clade assignment via VirtualPhenotype. 10th Conference on Retroviruses and Opportunistic Infections, 31 Feb-10-Mar, 2002, Boston, MA, Abstract 688
- Gonzales MJ, Wu TD, Taylor J, Belshkaya I, Kantor R, Israelski D, Chou S, Zolopa AR, Fessel WJ, Shaller RW. Extended spectrum of HIV-1 reverse transcriptase mutations in patients receiving multiple nucleoside analog inhibitors. AIDS 2003;17(6):791-799
- Benjamin Y and Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Stat Methodol. 1995; 57:289-300
- Fitch WM, Margoliash E. Construction of phylogenetic trees. Science 1967; 155(760):279-284
- Tsai K, Helek-Snyder G, Ewale F, Su C, Chan B, Chiu S, Sista P and Cammack N. Predicting phenotypic susceptibility levels of HIV reverse transcriptase and protease from genotype using tree modeling methods. 10th Conference on Retroviruses and Opportunistic Infections, 31 Feb-10-Mar, 2002, Boston, MA, Abstract 567
- Balzarini J, Plebanski H, Emond R, De Clercq E. A novel mutation (F227L) arises in the reverse transcriptase of human immunodeficiency virus type 1 on dose-escalating treatment of HIV type-1-infected cell cultures with the nucleoside reverse transcriptase inhibitor thiooxanilide UC-