Journal of NeuroVirology, 8: 281–294, 2002
 ISSN 1355–0284/02 $12.00+.00
DOI: 10.1080/13550280290100716

The reverse transcriptase sequence of human immunodeficiency virus type 1 is under positive evolutionary selection within the central nervous system

Kelly J Huang,[1] Gerald M Alter,[1] and Dawn P Wooley[1,2]

[1]*Department of Biochemistry and Molecular Biology, Wright State University School of Medicine, Dayton, Ohio, USA; and* [2]*Center for Retrovirus Research, The Ohio State University, Columbus, Ohio, USA*

The human immunodeficiency virus type 1 (HIV-1) enters the central nervous system (CNS) during the acute phase of infection and causes AIDS-related encephalitis and dementia in 30% of individuals. Previous studies show that HIV-1 sequences derived from the CNS of infected patients, including the sequence encoding reverse transcriptase (RT), are genetically distinct from sequences in other tissues. The hypothesis of the current study is that the RT sequence of HIV-1 is under positive selection within the CNS. Multiple alignments of non–CNS-derived and CNS-derived HIV-1 RT sequences were constructed using the ClustalW 1.8 program. The multiple alignments were analyzed with the Synonymous/Nonsynonymous Analysis Program. Codon positions 122–125, 135–149, and 166–212 of the CNS-derived RT sequences underwent a greater accumulation of nonsynonymous than synonymous substitutions, which was markedly different from the analysis results of the non–CNS-derived RT sequences. These residues are located in the finger and palm subdomains of the RT protein structure, which encodes the polymerase active site. The analysis of CNS-derived partial-length RT sequences that encompass these regions yielded similar results. A comparison of CNS-derived RT sequences to a non–CNS-derived RT consensus sequence revealed that a majority of the nonsynonymous substitutions resulted in a specific amino acid replacement. These results indicate that reverse transcriptase is under positive selection within the CNS. The amino acid replacements were visualized on a three-dimensional structure of HIV-1 RT using the Sybyl software suite. The protein structure analysis revealed that the amino acid replacements observed among the CNS-derived sequences occurred in areas of known structural and functional significance. *Journal of NeuroVirology* (2002) **8,** 281–294.

Keywords: retrovirus; brain; population genetics; protein structure

Introduction

The human immunodeficiency virus type 1 (HIV-1) enters the central nervous system (CNS) during the acute phase of infection (Shaw *et al*, 1985; Di Stefano *et al*, 1996; Strizki *et al*, 1996; Gisslen *et al*, 1997; An *et al*, 1999). During the later stages of disease progression, approximately 30% of HIV-1 seropositive patients experience neurological debilitation (McArthur, 1987; Masliah *et al*, 2000). Pathologies

Address correspondence to Dr Dawn P Wooley, Wright State University School of Medicine, Department of Biochemistry and Molecular Biology, 3640 Colonel Glenn Highway, Dayton, Ohio 45435, USA. E-mail: dawn.wooley@wright.edu

We thank Dr Virgilio Ponferrada, Dr Lisa Bircher, and Jeff Taylor for comments on the manuscript. The Biomedical Sciences PhD Program supported KJH. Public Health Service Grant CA72239 from the National Institutes of Health and a Research Challenge Grant from the state of Ohio supported this research.

Received 10 December 2001; revised 13 March 2002; accepted 18 April 2002.

associated with HIV-1 infection of the CNS include HIV encephalitis and HIV-associated dementia (McArthur *et al*, 1989; Glass *et al*, 1993). Despite the extensive use of antiretroviral drugs, HIV encephalitis remains a complication of AIDS (Masliah *et al*, 2000), underscoring the importance of understanding HIV-1 infection of the CNS.

The replicatingHIV-1 population in the CNS is compartmentalized by anatomical distance. The compartmentalization of the virus population is evident on both genotypic (Wong *et al*, 1997; van't Wout *et al*, 1998; Gatanaga *et al*, 1999; Venturi *et al*, 2000) and phenotypic (Gatanaga *et al*, 1999; Cunningham *et al*, 2000; Ellis *et al*, 2000; Haas *et al*, 2000; Venturi *et al*, 2000) levels. In terms of genotype, studies show that CNS-derived sequences and non–CNS-derived sequences, obtained from the same individual, cluster separately from each other in phylogenetic analyses (Wong *et al*, 1997). Additionally, in the presence of antiretroviral treatment, mutations conferring drug resistance evolve independently in the cerebrospinal fluid (CSF) as compared to peripheral tissues (Gatanaga *et al*, 1999; Cunningham *et al*, 2000; Venturi *et al*, 2000). In terms of phenotype, studies show that HIV-1-infected patients undergoing highly active antiretroviral therapy experience a slower regression of HIV-1 RNA levels in the CSF relative to the regression of plasma levels (Iftimovici *et al*, 1998; Ellis *et al*, 2000; Haas *et al*, 2000).

Compartmentalization within the CNS results in a virus population that is distinct from the peripheral virus population, and is designated as a quasispecies (Epstein *et al*, 1991; Korber *et al*, 1994b; Power *et al*, 1995; Wong *et al*, 1997; Bratanich *et al*, 1998; van't Wout *et al*, 1998; Gatanaga *et al*, 1999; Venturi *et al*, 2000). Viral quasispecies are reservoirs of phenotypic variants. For example, mutations associated with antiretroviral drug resistance are found in the CSF of drug-naïve patients (Najera *et al*, 1995). Unique selective pressures provided by the microenvironment of the CNS contribute to the quasispecies nature of HIV-1 within the CNS and include local immune response, constraints on viral entry and replication rates due to target cell availability, as well as local availability of antiretroviral drugs. These selective pressures result in the emergence of HIV-1 CNS-derived quasispecies, evidenced by the sequence analysis of four regions of the HIV-1 genome including the long terminal repeat (LTR), *env*, *tat*, and reverse transcriptase (RT) (Power *et al*, 1995; Strizki *et al*, 1996; Corboy and Garl, 1997; Wong *et al*, 1997; Bratanich *et al*, 1998; Krebs *et al*, 1998; van't Wout *et al*, 1998).

With regard to pathology, it is important to determine whether these viral coding sequences undergo positive evolutionary selection within the CNS. Positive selection is evident by prominent protein sequence changes within a population. Protein sequence changes can occur when the replicating viral genome undergoes nucleotide substitution, which is a frequent event during HIV-1 replication (Mansky and Temin, 1995). Recent studies show that the HIV-1 LTR, *env*, and *tat* sequences are under positive selective pressure within the CNS (Corboy and Garl, 1997; Bratanich *et al*, 1998; Krebs *et al*, 1998; van't Wout *et al*, 1998). However, studies regarding whether HIV-1 RT is under positive selection within the CNS are conflicting and inconclusive. A phylogenetic analysis by Wong *et al* (1997) reveals the *in vivo* compartmentalization and quasispecies nature of RT within the CNS by comparing CNS-derived sequences and peripheral sequences. However, the study included only four patient sequences and did not specifically address the issue of positive selection. Bratanich *et al* (1998) report that RT is not under positive selection within the CNS as determined by the comparison of sequences from patients with and without HIV-1-associated dementia. This study addresses pathological distinction within the CNS; however, the study does not address whether the RT sequence is under positive selection within the CNS, as would be evident by comparing CNS-derived RT sequences with non–CNS-derived RT sequences.

The selective pressure of antiretroviral drugs influences the representative RT sequences within a population, and may alter analyses for tissue-specific reverse transcriptase sequences. Studies to date include RT sequences derived from the CNS of patients treated with antiretroviral drugs that target reverse transcriptase (Wong *et al*, 1997; Bratanich *et al*, 1998; Gatanaga *et al*, 1999; Venturi *et al*, 2000). It is clear that an in-depth analysis for tissue-specific selection of the HIV-1 RT sequence derived from the CNS is needed. Therefore, the aim of this study was to determine if HIV-1 RT is under positive selection within the CNS, and if so, to identify structural correlations of the protein sequence alterations.

The hypothesis that HIV-1 RT is under positive selection within the CNS was tested by the comparison of nucleotide substitution among CNS-derived sequences and non–CNS-derived sequences. There are two types of nucleotide substitutions, synonymous and nonsynonymous. A synonymous substitution is a nucleotide base change that does not result in an amino acid replacement and therefore does not alter the protein sequence. A nonsynonymous substitution is a nucleotide base change that results in an amino acid replacement and therefore alters the protein sequence. Amino acid replacements within a protein sequence are subject to selection. Positive selection is characterized by higher rates of nonsynonymous than synonymous substitution, termed substitution behavior.

This study compared the substitution behavior of RT sequences derived from the CNS to RT sequences derived from peripheral tissue. Multiple sequence alignments were constructed using the ClustalW 1.8 program (Jeanmougin *et al*, 1998). The Synonymous/Nonsynonymous Analysis

Program (SNAP) (Korber, 1994a) was used to determine substitution behavior. In addition, the Sybyl software suite was used to visualize protein sequence changes among CNS-derived sequences on a three-dimensional level. The coordinates derived from a crystal structure of HIV-1 RT (Huang *et al*, 1998) were used to identify structural correlations. The results of this study indicate that HIV-1 RT is under positive selection within the CNS, represented by a greater accumulation of nonsynonymous than synonymous substitutions among CNS-derived RT sequences, which was markedly different from the analysis results for the non–CNS-derived RT sequences. Specifically, the finger and palm subdomains of RT underwent more nonsynonymous than synonymous substitution among CNS-derived sequences. The finger and palm subdomains encode the polymerase active site of RT; therefore, the locations of the amino acid replacements within the tertiary structure of RT were determined. The protein structure analysis revealed that the amino acid replacements occur in known regions of structural and functional significance.

Results

Positive selection is operating on HIV-1 RT in the CNS

The SNAP analysis of full-length non–CNS-derived HIV-1 RT sequences is plotted in a graph as cumulative substitution behavior (Figure 1A). There were more synonymous substitutions than nonsynonymous substitutions throughout the non–CNS-derived RT sequences. The sequences are highly similar, as all sequences were derived from the peripheral blood mononuclear cells (PBMCs) or plasma of antiretroviral drug-naïve patients infected with HIV-1 clade B of group M. The RT sequences classified as clade B of group M were used in this study due to the greater availability of such sequences in the database. This result provided a baseline for the amount of nonsynonymous substitution relative to the amount of synonymous substitution to be expected in a comparison of RT sequences that are highly similar, and not under positive selection.

A multiple alignment was constructed using full-length RT sequences derived from the CNS of antiretroviral drug-naïve patients. The sequences derived from the CNS are classified as HIV-1 clade B of group M. The cumulative substitution behavior was plotted in a graph using SNAP (Figure 1B). Codon positions 122–125, 135–149, and 166–212 of the CNS-derived RT sequences accumulated more nonsynonymous than synonymous substitutions, which was markedly different from the analysis results of the non–CNS-derived RT sequences. This was due to an increased rate of nonsynonymous substitution relative to synonymous substitution among CNS-derived sequences in these regions. These regions of the RT protein sequence are located within the finger and palm subdomains of the RT protein structure. The RT protein structure is a heterodimer consisting of monomers p66 and p51. The monomers fold into a typical polymerase hand conformation consisting of subdomains referred to as the finger, palm, connection, and thumb (Kohlstaedt *et al*, 1992). Notably, the finger and palm subdomains encode the polymerase active site of the enzyme.

A greater number of partial-length CNS-derived RT sequences provided a more in-depth analysis of these regions of the RT sequence. The sequences used to create the plot of full-length non–CNS-derived RT sequences (Figure 1A) were used to create a partial-length (the first 219 codons of RT) plot of non–CNS-derived RT sequences shown in Figure 2A. A multiple alignment of the first 219 codons of RT sequences derived from the CNS was constructed. The sequences derived from the CNS are classified as HIV-1 clade B of group M. The low availability of viral sequences derived from antiretroviral drug-naïve patients necessitated including partial RT sequences from individuals receiving antiretroviral therapy. The cumulative substitution behavior was plotted in a graph using SNAP (Figure 2B). There were several regions within codon positions 1–219 of CNS-derived RT sequences that underwent a greater rate of nonsynonymous than synonymous substitution.

The partial CNS-derived RT sequence analysis included sequences from a study by Wildemann *et al* (1993). Partial RT sequences recovered from the plasma and CSF of the three patients in this study were available in the database. Therefore, it was possible to compare paired non–CNS- and CNS-derived sequences. The analysis showed similar trends as the full-length and partial-length RT sequence analyses (data not shown). Noteworthy, the trend was detected in an analysis of three sequences ($N = 3$), which substantiates the relevance of the full-length CNS-derived RT sequence analysis with an $N = 3$ (Figure 1).

It is important to note that the scale of the Y-axis is different for Figures 1A and 1B. More substitutions occurred among non–CNS-derived RT sequences as compared to CNS-derived reverse transcriptase sequences (Figures 1 and 2). These results are consistent with previous studies that report that the HIV-1 population within the CNS is genetically more homogenous than the virus population in peripheral tissues (Pang *et al*, 1991; Ball *et al*, 1994; Korber *et al*, 1994b; Di Stefano *et al*, 1996; van't Wout *et al*, 1998; Morris *et al*, 1999).

Characterization of HIV-1 CNS-derived RT sequence changes

The SNAP codon-by-codon analysis of substitution behavior was used as a guide to identify the codon positions in HIV-1 RT sequences derived from the CNS that underwent nonsynonymous substitutions

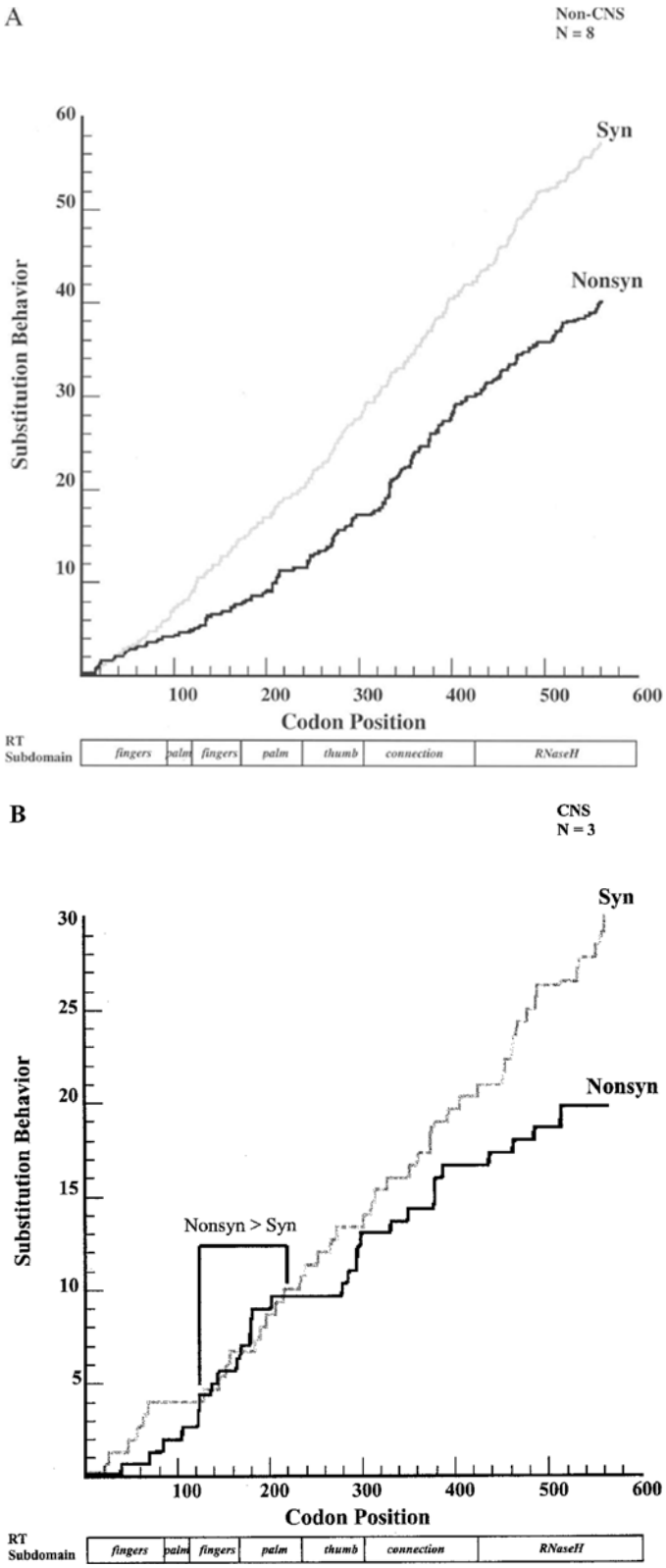


Figure 1 Substitution behavior of full-length non–CNS- and CNS-derived HIV-1 RT sequences. (**A**) A SNAP plot of the cumulative substitution behavior among full-length HIV-1 RT sequences derived from the PBMCs of antiretroviral drug-naïve patients infected with HIV-1 clade B of group M. (**B**) A SNAP plot of the cumulative substitution behavior among full-length RT sequences derived from the CNS of antiretroviral drug-naïve patients infected with HIV-1 clade B of group M. The bracket above the plot indicates regions where more nonsynonymous substitution than synonymous substitution occurred among RT sequences derived from the CNS. The corresponding structural subdomains of RT are indicated below the codon position axis. Synonymous substitution = Syn; Nonsynonymous substitution = Nonsyn.

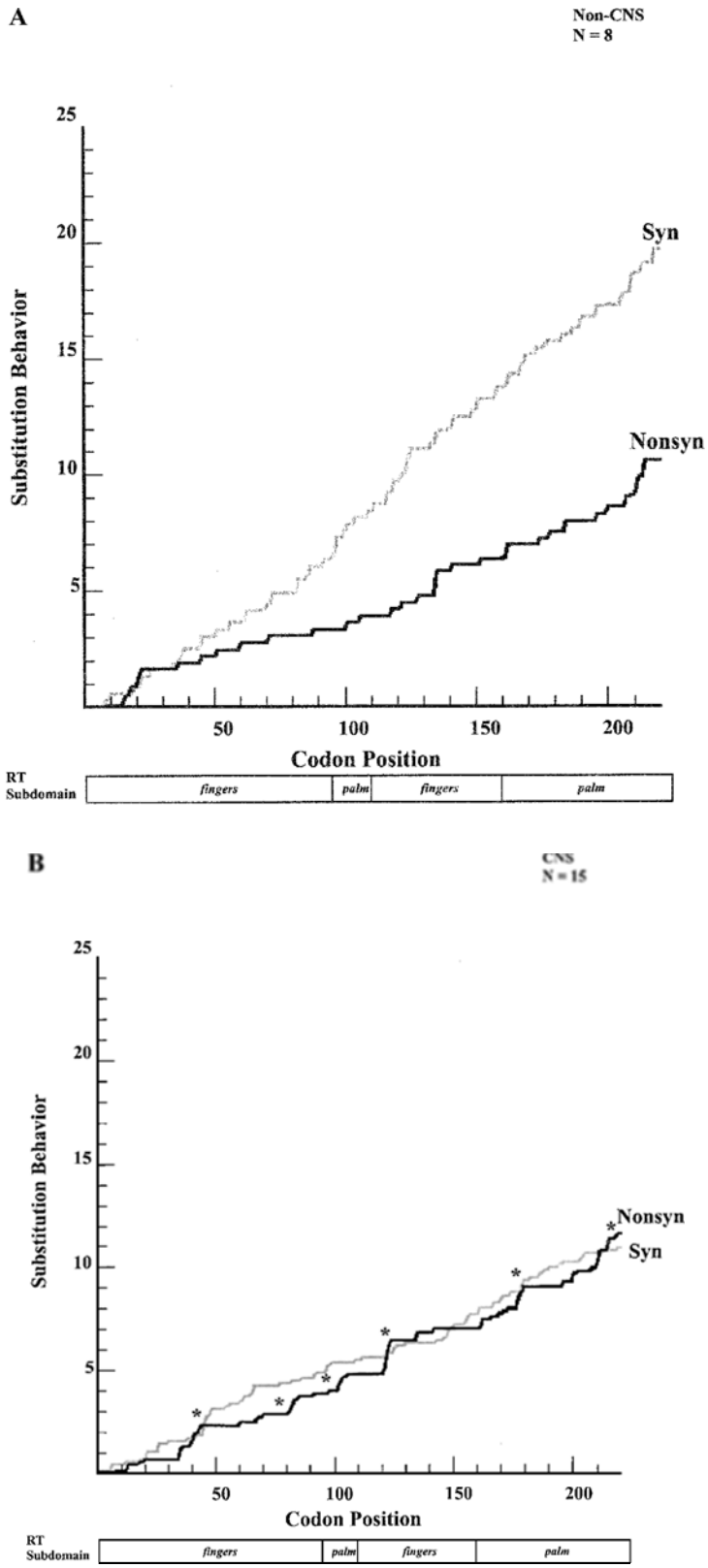


Figure 2 Substitution behavior of partial-length non–CNS- and CNS-derived HIV-1 RT sequences. (**A**) A SNAP plot of the cumulative substitution behavior of the first 219 codons of RT sequences derived from the PBMCs of antiretroviral drug-naïve patients infected with HIV-1 clade B of group M. (**B**) A SNAP plot of the cumulative substitution behavior of the first 219 codons of RT sequences derived from the CNS of patients infected with HIV-1 clade B of group M. The * indicates regions that underwent a greater rate of nonsynonymous substitution than synonymous substitution.

relative to an RT consensus sequence representative of non–CNS-derived RT sequences. This primary sequence data analysis revealed the specific amino acid replacements and potential structure-function correlations of the nonsynonymous substitutions that occurred among CNS-derived RT sequences. It was observed that the majority of nonsynonymous substitutions that occurred among CNS-derived RT sequences resulted in a single, specific amino acid replacement at a given codon position rather than a variety of possible amino acids. In addition, the amino acid replacements occurred in regions of the RT structure implicated in catalysis and fidelity. These results are summarized in Table 1.

The protein database coordinates derived from a crystal structure of HIV-1 RT (Huang *et al*,

Table 1 Amino acid replacements in the finger and palm subdomains of CNS-derived RT sequences

Codon position	*Consensus sequence*	*CNS sequence*	*Color in Figures 3 and 4*	*Predicted functional correlation*
8	Val	Ile	Blue	
13	Lys	Arg	Magenta	
17	Asp	Gly	Blue	
20	Lys	Arg	Blue	Template grip[3]
35	Val	Leu, Arg, Ile	Blue	
36	Glu	Asp	Blue	
39	Thr	Ile	Red	
40	Glu	Val	Red	
41	Met	Leu	Red	Drug resistant[6]
43	Lys	Glu	Red	
44	Glu	Ala	Red	
60	Val	Ile	Blue	Stabilizing[3] and substrate binding[4]
67	Asp	Asn	Yellow	Drug resistant[6] and substrate binding[2,4]
68	Ser	Gly	Yellow	Substrate binding[3,4]
70	Lys	Arg	Yellow	Drug resistant[6] and substrate binding[3]
81	Asn	Tyr	Magenta	
82	Lys	Arg	Magenta	
83	Arg	Lys	Magenta	
85	Gln	Leu	Magenta	
91	Gln	Pro	Blue	Template grip[1] and substrate binding[3]
98	Ala	Ser	Blue	Drug resistant[6]
102	Lys	Gln	Orange	Increased RT activity[5]
103	Lys	Arg, Thr	Orange	Drug resistant[6] and wild-type RT activity (Thr)[5]
104	Lys	Asn	Orange	
106	Val	Ile	Orange	Drug resistant[6] and wild-type RT activity[5]
121	Asp	His, Tyr	Blue-green	
122	Lys	Glu, Gln, Pro	Blue-green	
123	Asp	Asn, Glu	Blue-green	Wild-type RT activity (Asn, Glu)[5]
124	Phe	Asn	Blue-green	
135	Ile	Thr, Val	Blue	Wild-type RT activity[5]
142	Ile	Thr	Blue	
162	Ser	Cys	Blue	Decreased RT activity[5] and substrate binding[2]
166	Lys	Thr	Blue	Decreased RT activity[5]
169	Glu	Asp	Blue	Increased RT activity[5]
171	Phe	Tyr	Blue	Wild-type RT activity[5]
173	Lys	Arg	Blue	Increased RT activity[5]
177	Asp	Glu	Orange	Increased RT activity[5]
178	Ile	Met, Leu	Orange	Increased RT activity (Leu)[5]
179	Val	Ile	Orange	Drug resistant[7] and increased RT activity[5]
196	Gly	Glu	Blue	
200	Thr	Ala, Ile	Blue	Decreased RT activity (Ile) and increased RT activity (Ala)[5]
202	Ile	Val	Blue	Increased RT activity[5]
208	His	Tyr	Blue	
210	Leu	Trp	Cyan	Drug resistant[6]
211	Lys	Arg	Cyan	
214	Phe	Leu	Cyan	
215	Thr	Tyr	Cyan	Drug resistant[6]
218	Asp	Glu	Yellow	Substrate binding[1]
219	Lys	Gln	Yellow	Drug resistant[6] and substrate binding site[1]

[1]Ding *et al*, 1997.
[2]Halvas *et al*, 2000.
[3]Sarafinos *et al*, 1999.
[4]Shah *et al*, 2000.
[5]Wrobel *et al*, 1998.
[6]Shafer *et al*, 1999.

1998, protein database number 1RTD) were used to visualize the amino acid replacements that occurred among CNS-derived sequences on a three-dimensional structure of the RT p66 monomer with bound primer template and dTTP (Figure 3). The majority of the amino acid replacements were placed in color-coded groups, based on proximity to each other, within the structure of the RT p66 monomer. The groups were located on all sides of the finger and palm subdomains relative to the bound dTTP.

For clarity, close-up views of each group of amino acid side chains are shown in relation to the bound dTTP in Figure 4. The red group of amino acid replacements (Figure 4A) is part of an alpha helix. The orange group of amino acid replacements (Figure 4B) was located on the external protein surface. The yellow group of amino acid replacements in Figure 4C associated with the dNTP binding site (Ding *et al*, 1997; Sarafianos *et al*, 1999) as well as regions thought to modulate the fidelity of reverse transcriptase (Shah *et al*, 2000). The blue-green group (Figure 4D) was located on the external surface of the protein. The cyan group of amino acid replacements (Figure 4E) have not been reported to be involved in dNTP binding; however, the residues are close to the bound dTTP. The magenta group (Figure 4F) was located on the external surface of the protein, and is near residues involved in fidelity (Halvas *et al*, 2000).

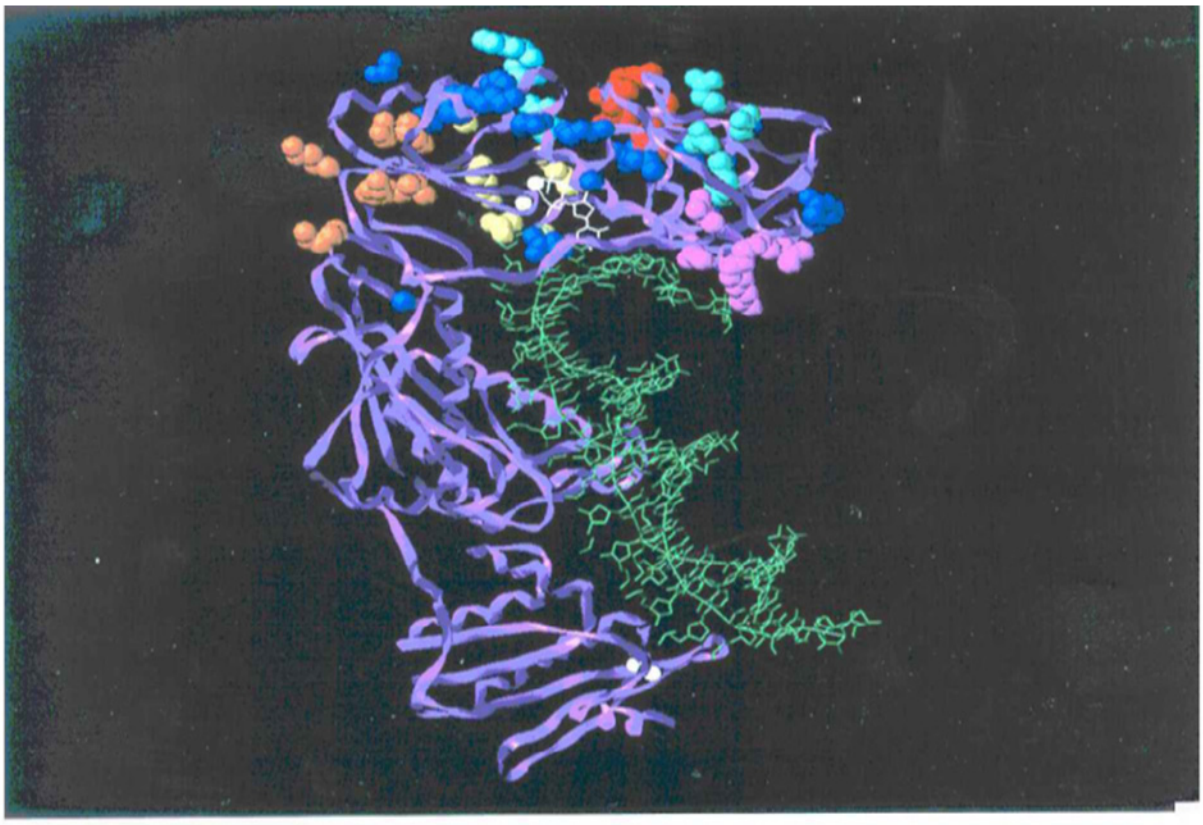
Discussion

The results of this study support the hypothesis that HIV-1 RT is under positive selection within the CNS. A comparison of the substitution behavior of non–CNS- and CNS-derived full-length RT sequences (Figure 1) revealed a difference in the rates and accumulation of nonsynonymous substitution relative to synonymous substitution in these sequences. The non–CNS-derived RT sequences underwent synonymous substitution more frequently than nonsynonymous substitution. These results indicate that RT is not undergoing frequent amino acid replacement within peripheral tissues, and the protein sequence is optimal in terms of function in the periphery. In contrast, the CNS-derived RT sequences underwent nonsynonymous substitution more frequently relative to synonymous substitution in the regions spanning codon positions 122–125, 135–149, and 166–212. These results indicate that RT is not optimal in terms of function in these regions within the CNS, and is undergoing positive selection.

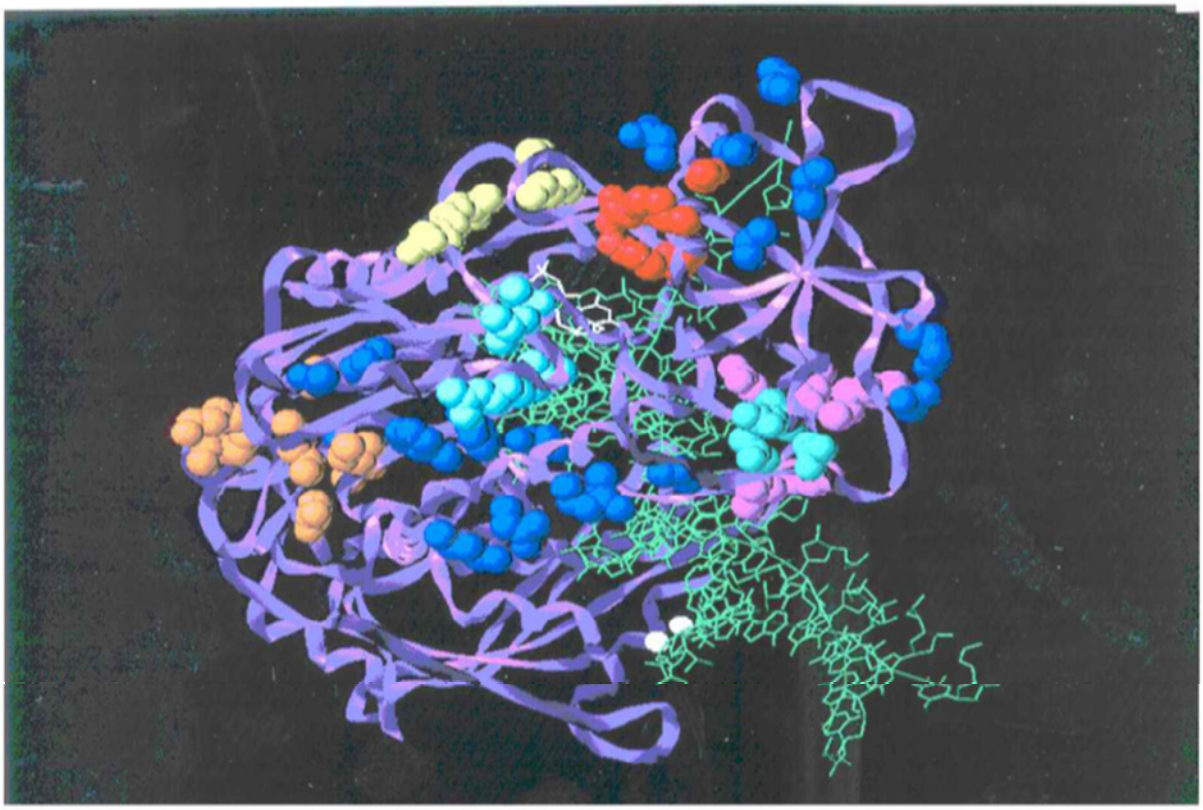
This is the first report to show that the RT protein sequence changes among CNS-derived sequences fall within the finger and palm subdomains of the RT protein structure. The region spanning codon positions 122–212 of the RT structure comprise those subdomains, which encode the polymerase active site of the enzyme. The substitution behavior analysis of CNS-derived partial-length RT sequences (Figure 2) confirmed that distinct changes within the finger and palm subdomains of RT occur among HIV-1 CNS-derived sequences, and further demonstrates that positive selection is operating on the HIV-1 reverse transcriptase sequence within the CNS. This finding is important, as a previous study dismisses the significance of CNS-derived RT sequences based on the inconsistency of specific sequence alterations among CNS-derived RT sequences (Bratanich *et al*, 1998).

There are two significant differences between the non–CNS- and CNS-derived sequences used in this study. First, the CNS-derived partial-length RT sequence analysis included sequences recovered from patients receiving antiretroviral therapy, possibly altering the results of this analysis. However, the CNS-derived partial-length RT sequence analysis yielded similar trends to the analysis of CNS-derived full-length RT sequences obtained from antiretroviral drug-naïve patients. Second, the majority of the CNS-derived partial-length RT sequences were recovered from RNA, whereas the majority of the non–CNS-derived partial-length RT sequences were recovered from DNA. To resolve this difference, it should be noted that the CNS-derived full-length RT sequences were recovered from DNA and yielded similar trends to the CNS-derived partial-length RT sequence analysis, which contends that the partial non–CNS- and CNS-derived comparisons are appropriate (Table 2).

As previously noted, the homogeneity among CNS-derived HIV-1 sequences observed in this study is consistent with previous reports (Di Stefano *et al*, 1996; van't Wout *et al*, 1998; Morris *et al*, 1999). The homogeneity of CNS-derived HIV-1 quasispecies may be the product of a lower viral replication rate within the CNS, as a lower viral replication rate decreases the opportunity for nucleotide substitution. If this was accurate, then the substitution behavior patterns of non–CNS- and CNS-derived sequences plotted by SNAP should look similar, despite a decreased opportunity for nucleotide substitution. The results of this study discount the possibility of a lower viral replication rate due to the difference in substitution behavior of non–CNS- and CNS-derived sequences (Figures 1 and 2). In addition, it is important to note that the majority of the CNS-derived sequences were recovered from RNA (Table 2), which implies active viral replication. Recent studies indicate that viral replication within the CNS may not be significantly lower from that in peripheral tissues (Stankoff *et al*, 1999). Alternative explanations for the homogeneity among HIV-1 sequences derived from the CNS could be highly selective constraint, or a decrease in the mutation rate during viral replication. A decrease in the mutation rate during viral replication is relevant to this study, as sequence variation within the finger and palm subdomains of RT alter the rate and pattern of mutation during reverse transcription (Kim *et al*, 1999; Lewis *et al*, 1999; Boyer and Hughes, 2000;



A



B

Figure 3 Images from the Sybyl software-assisted visualization of the amino acid replacements that occurred among CNS-derived RT sequences. The image shows the p66 monomer of HIV-1 RT with bound primer/temple (green), dTTP (white), Mg^{++} cations (white), and the location of the amino acid replacements that occurred among CNS-derived sequences. (**A**) A side view. (**B**) Top view of the structure shown in panel A. The amino acid replacements were divided into groups based on proximity to each other and color-coded as red, orange, yellow, green-blue, cyan, and magenta. The dark blue residues are nongrouped amino acid replacements.

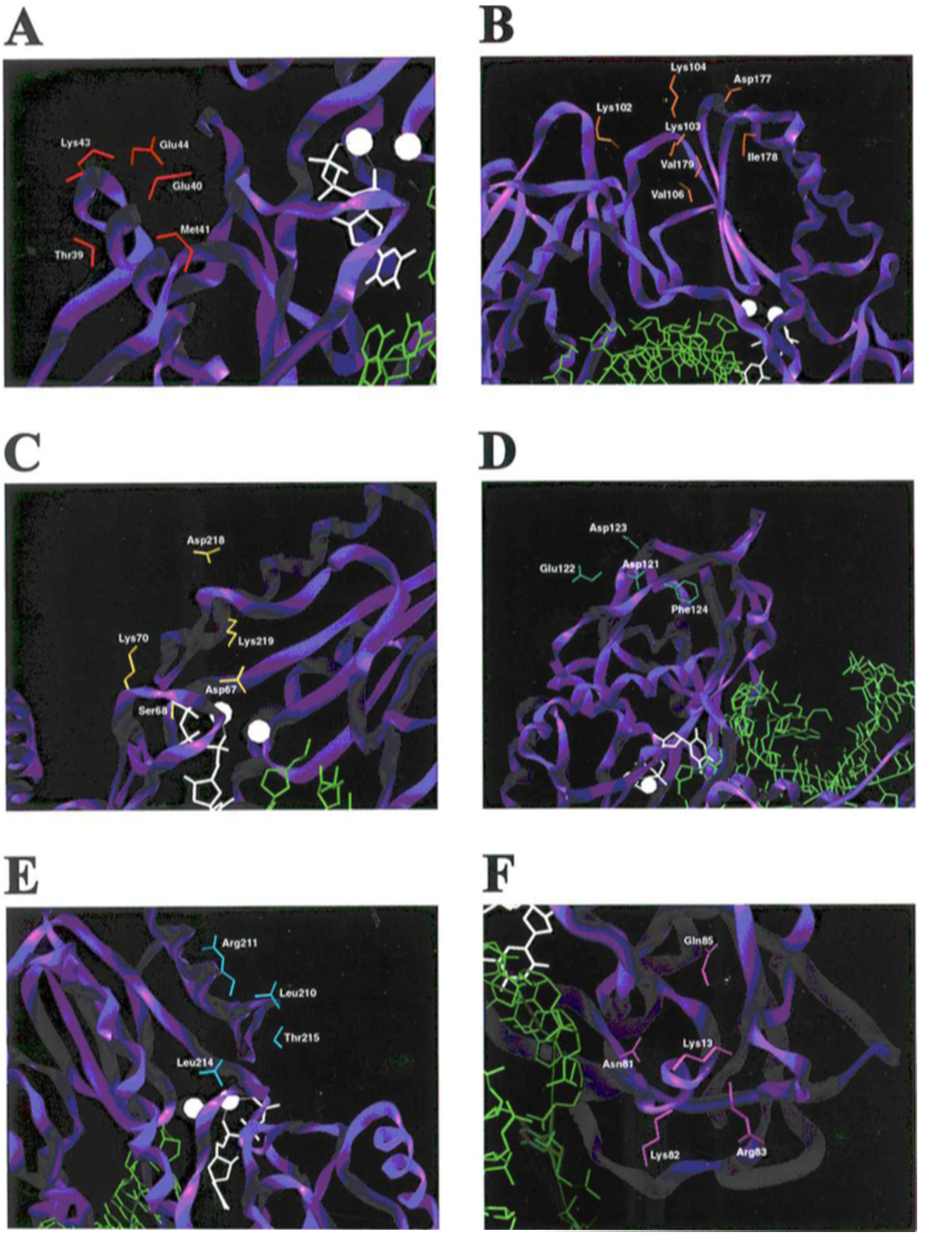


Figure 4 Close-up views of each of the grouped amino acid replacements. (**A**) Represents the red group, which is part of an alpha helix. (**B**) Represents the orange group, which is located externally. (**C**) Represents the yellow group, which includes dNTP-binding sites. (**D**) Represents the blue-green group, which is located externally. (**E**) Represents the cyan group, which is located near the bound dTTP. (**F**) Represents the magenta group, which is located near residues involved in fidelity. The side chains of the amino acids are annotated and represent the consensus sequences (except the blue-green residue Phe124).

Table 2 Sources of HIV-1 sequences

Accession no	*Submitting author(s)*	*Biological source*
	HIV-1 peripheral reverse transcriptase sequences	
AY037268	Carr *et al*, 2001	DNA from HIV-1-seropositive patient PBMCs
AY037269	Carr *et al*, 2001	DNA from HIV-1-seropositive patient PBMCs
AY037270 (*)	Carr *et al*, 2001	DNA from HIV-1-seropositive patient PBMCs
AY037274	Carr *et al*, 2001	DNA from HIV-1-seropositive patient PBMCs
AY037282	Carr *et al*, 2001	DNA from HIV-1-seropositive patient PBMCs
U69584	Matala, 1996	RNA from HIV-1-seropositive patient plasma
AF331192	Lukashov *et al*, 2000	RNA from HIV-1-seropositive patient plasma
U34603	Guillon *et al*, 1995	DNA from HIV-1-seropositive patient PBMCs
	Full-length HIV-1 CNS reverse transcriptase sequences	
M93258	Li *et al*, 1992	DNA from HIV-1-seropositive patient brain tissue
U63632	O'Brien *et al*, 1991	DNA from HIV-1-seropositive patient brain tissue
AF042101	Oelrichs *et al*, 1998	DNA from HIRT supernatant of HIV-1 seropositive patient CSF after low passage with donor PBMCs
	Partial-length HIV-1 CNS reverse transcriptase sequences	
U63632	O'Brien *et al*, 1991	DNA from HIV-1-seropositive patient brain tissue
U82089 (+)	Lui *et al*, 1996	RNA from HIV-1-seropositive patient brain tissue
U82091 (+)	Lui *et al*, 1996	RNA from HIV-1-seropositive patient brain tissue
U82092 (+)	Lui *et al*, 1996	RNA from HIV-1-seropositive patient brain tissue
U82093 (+)	Lui *et al*, 1996	RNA from HIV-1-seropositive patient brain tissue
U82094 (+)	Lui *et al*, 1996	RNA from HIV-1-seropositive patient brain tissue
U82095 (+)	Lui *et al*, 1996	RNA from HIV-1-seropositive patient brain tissue
U82096 (+)	Lui *et al*, 1996	RNA from HIV-1-seropositive patient brain tissue
U82097 (+)	Lui *et al*, 1996	RNA from HIV-1-seropositive patient brain tissue
U82098 (+)	Lui *et al*, 1996	RNA from HIV-1-seropositive patient brain tissue
X70589 (+)	Haas, 1993	DNA from HIV-1-seropositive patient CSF
X70603 (+)	Haas, 1993	DNA from HIV-1-seropositive patient CSF
X70634 (+)	Haas, 1993	DNA from HIV-1-seropositive patient CSF
M93258	Li *et al*, 1992	DNA from HIV-1-seropositive patient brain tissue
AF042101	Oelrichs *et al*, 1998	DNA from HIRT supernatant of HIV-1 seropositive patient CSF after low passage with donor PBMCs

All sequences are classified as HIV-1 group M clade B. Reverse transcriptase sequences that were isolated from patients receiving antiretroviral drugs (+) or unknown antiretroviral drug use (*) are indicated.

Cases-Gonzalez *et al*, 2000; Jonckheere *et al*, 2000; Kaushik *et al*, 2000; Shah *et al*, 2000; Weiss *et al*, 2000).

The codon-by-codon account of substitution behavior from SNAP was used to identify the residues that underwent amino acid replacement within the CNS relative to a non–CNS-derived consensus sequence (Table 1). It was observed that the majority of nonsynonymous substitutions among CNS-derived RT sequences resulted in a single, specific amino acid replacement at a given codon position rather than a variety of possible amino acids. This observation supports the hypothesis that positive selection, a directive force for a specific sequence, is operating on HIV-1 RT in the CNS. There were eight amino acid replacements among the CNS-derived sequences that are reported to increase the activity of the enzyme (Wrobel *et al*, 1998). An increase in activity of the enzyme is a phenotype upon which positive selection may operate. Also, there were five amino acid replacements among the CNS-derived sequences that maintain wild-type activity (Wrobel *et al*, 1998). Typically, wild-type activity is a phenotype that would not be susceptible to selection. However, the amino acid replacements that result in wild-type activity in these sequences may give rise to alternative phenotypes. For example, Lys103Thr (Figure 4B) maintains wild-type activity, but is also a drug resistance phenotype susceptible to positive selection (Shafer *et al*, 1999). There were only three amino acid replacements among the CNS-derived sequences that decrease activity (Wrobel *et al*, 1998). The primary structure analysis resulted in the identification of residues potentially involved in the catalytic activity and substrate fidelity of the enzyme as well as antiretroviral drug resistant phenotypes.

The Sybyl software suite was used to visualize the structural correlations of the amino acid replacements that occurred among CNS-derived sequences on a three-dimensional structure of the RT p66 monomer with bound template and dTTP, based on the protein database coordinates of a crystal structure of HIV-1 RT (Huang *et al*, 1998) (Figures 3 and 4). Amino acid replacements were commonly located close to dTTP binding sites and template grip sites within the RT structure. The regions within the finger and palm subdomains that underwent a greater rate of nonsynonymous than synonymous substitution (Figure 2B) correlate with the color-coded groups of amino acid replacements shown in Figure 4.

Functional consequence of amino acid replacements can be suggested according to their grouping in Figures 3 and 4. The red group may be involved in dNTP-binding as Glu40Val and Glu44Ala face the dNTP-binding site, and resulted in the replacement of negatively charged amino acids with noncharged amino acids (Figure 4A). Moreover, the Met41Leu (red) results in decreased sensitivity to AZT (Shafer *et al*, 1999). All the amino acid replacements of the orange group in Figure 4B with the exception of Lys104Asn are shown to either maintain or increase the activity of the enzyme (Wrobel *et al*, 1998). The yellow group of amino acid replacements in Figure 4C associated with the dNTP binding site (Ding *et al*, 1997; Sarafianos *et al*, 1999) as well as regions thought to modulate the fidelity of reverse transcriptase (Shah *et al*, 2000). In fact, residue 219 (yellow) is a dNTP-binding residue, and Lys219Gln (yellow) results in resistance to nucleoside RT inhibitors (NRIs). The amino acid replacement Asp121Tyr (blue-green) increases activity, and the amino acid replacement Asp123Asn or Glu (blue-green) maintains activity (Figure 4D). The amino acid replacements shown in cyan (Figure 4E) are not reported to be involved in dNTP binding, however the residues appear close to the bound dTTP in the three-dimensional structure. Furthermore, the amino acid replacements Leu210Trp and Thr215Tyr (cyan) result in resistance to NRIs. The magenta group (Figure 4F) was located on the external surface of the protein, and is near residues involved in fidelity (Halvas *et al*, 2000). Aside from the amino acid replacements that increase activity, the nongrouped amino acid replacements (dark-blue) were farther away from the dTTP-binding site and located on the external surface of the protein.

The principle conclusion drawn from this study is that the finger and palm subdomains of the HIV-1 RT sequence are under positive selection in the CNS. There are several possible reasons for RT to be under positive selection within the CNS. First, the CNS is predominated by nondividing cells that may offer a unique environment in terms of polymerase activity including lower dNTP pools (Julias and Pathak, 1998) and other replication machinery components. Second, there may be CNS-specific cellular factors that inhibit reverse transcription. Third, the immune response in the CNS is different than that of the periphery, possibly impacting reverse transcriptase activity. Finally, there are a lower number of cellular targets within the CNS that provide a productive infection, possibly resulting in a highly fit replicating viral population.

The protein data analysis in this study provides a guide for future studies concerning structure-function relationships of HIV-1 RT. Future studies should address the phenotypic analysis of CNS-derived HIV-1 RT sequences and provide a greater understanding of the significance of CNS-specific reverse transcriptase sequences.

Materials and methods

Sequences

The sequences used in this study were obtained from the Los Alamos HIV Sequence Database ⟨http://hiv-web.lanl.gov/⟩ and are listed in Table 2. All of the sequences used in this study are classified as HIV-1 clade B of group M. HIV-1 is divided into three groups main (M), outlier (O), and new (N). The M group includes the subtypes, or clades, of HIV-1 that are most prominent worldwide. The RT sequences classified as clade B of group M were used in this study due to the greater availability of such sequences in the database. In addition, the sequences used in this study were obtained directly from HIV-1-seropositive patient tissue, with the exception of one CNS-derived sequence (AF042101). Therefore, the sequences are representative of the replicating virus population within a host, and are not a product of cell culture. This criterion greatly reduced the number sequences available in the database that could be included in this study. The CNS-derived sequence AF042101 was included to increase the number of full-length CNS-derived sequence comparisons. An important caveat is that although the open reading frames of the sequences included in this study are intact, it cannot be ruled out that these sequences are from replication incompetent viral genomes.

Substitution behavior analysis of non–CNS-derived HIV-1 RT sequences

A multiple alignment of six non–CNS-derived RT sequences was constructed using the ClustalW 1.8 program (Jeanmougin *et al*, 1998). ClustalW 1.8 performs a comparison of all sequences to each other (pairwise alignments), groups sequences by similarity, and the output is controlled for gaps in the final multiple alignment. The non-CNS RT sequences were derived from the PBMCs or plasma of antiretroviral drug-naïve patients. The multiple alignment of non–CNS-derived RT sequences was submitted to SNAP (Korber, 1994a) to analyze the substitution behavior, defined as the amount of synonymous substitution relative to the amount of nonsynonymous substitution observed between compared DNA sequences. SNAP is available on the Los Alamos HIV sequence database website ⟨http://hiv-web.lanl.gov/⟩, and has been used previously in HIV-1 research (Ganeshan *et al*, 1997). The SNAP program calculates the synonymous and nonsynonymous substitution rates of all pair-wise comparisons in a multiple alignment of codon-aligned DNA sequences based on the method of Nei and Gojobori (1986), which incorporates a statistic of Ota and Nei (Ota and Nei, 1994). The SNAP plot output represents the most common calculations for all pair-wise comparisons. Changes in substitution behavior are indicative of selective forces operating within a sequence.

Substitution behavior analysis of CNS-derived HIV-1 RT sequences

Multiple sequence alignments of full-length and partial-length HIV-1 RT sequences derived from the CNS of patients infected by HIV-1 clade B of group M were constructed using ClustalW 1.8 (Jeanmougin *et al*, 1998). Three full-length RT sequences derived from the CNS of antiretroviral drug-naïve patients were available on the Los Alamos HIV sequence database. In addition, there were 15 partial-length RT sequences derived from the CNS of infected patients available on the database. The partial sequence alignment included 12 sequences derived from patients receiving antiretroviral drugs, due to the low availability of sequences derived from drug-naïve patients. The partial sequence alignment included the first 219 codons of the RT sequence. The multiple alignments were submitted to SNAP for substitution behavior analysis.

Characterization of HIV-1 CNS-derived RT sequence changes

A SNAP analysis includes a codon-by-codon account of substitution behavior, in addition to the cumulative substitution behavior output in graph format, that was used to identify the codon positions within the CNS-derived RT sequences that undergo nonsynonymous substitution relative to a non–CNS-derived RT consensus sequence. The RT consensus sequence was created using the same 6 RT sequences used to analyze substitution behavior in non–CNS-derived RT sequences. The predominant sequence pattern among the six sequences was determined from a multiple alignment and directly corresponds to the consensus sequence. If a site within the multiple alignment showed an equal frequency of two different bases, one of the two possible bases was assigned at random. A multiple alignment was constructed that included CNS-derived RT sequences and the consensus RT sequence. The primary structure data obtained from this analysis was used to predict structural and/or functional correlations.

The Sybyl software suite was used to visualize the structural relationship of the protein sequence changes among CNS-derived RT sequences based on the crystal structure coordinates of HIV-1 RT with bound primer/template and dTTP (protein database number 1RTD) (Huang *et al*, 1998).

References

An SF, Groves M, Gray F, Scaravilli F (1999). Early entry and widespread cellular involvement of HIV-1 DNA in brains of HIV-1 positive asymptomatic individuals. *J Neuropathol Exp Neurol* **58:** 1156–1162.

Ball JK, Holmes EC, Whitwell H, Desselberger U (1994). Genomic variation of human immunodeficiency virus type 1 (HIV-1): Molecular analyses of HIV-1 in sequential blood samples and various organs obtained at autopsy. *J Gen Virol* **75:** 67–79.

Boyer PL, Hughes SH (2000). Effects of amino acid substitutions at position 115 on the fidelity of human immunodeficiency virus type 1 reverse transcriptase. *J Virol* **74:** 6494–6500.

Bratanich AC, Liu C, McArthur JC, Fudyk T, Glass JD, Mittoo S, Klassen GA, Power C (1998). Brain-derived HIV-1 tat sequences from AIDS patients with dementia show increased molecular heterogeneity. *J NeuroVirol* **4:** 387–393.

Carr J, Avila M, Gomez-Carrillo M, Salomon H, Hierholzer J, Watanaveeradej V, Pando M, Negrete M, Russell K, Russell KL, Sanchez, J, Birx D, Birx DL, Andrade R, Vinoles J, McCutchan F, McCutchan FE (2001). [online] ⟨http://hiv-web.lanl.gov/⟩

Cases-Gonzalez CE, Gutierrez-Rivas M, Menendez-Arias L (2000). Coupling ribose selection to fidelity of DNA synthesis. The role of Tyr-115 of human immunodeficiency virus type 1 reverse transcriptase. *J Biol Chem* **275:** 19759–19767.

Corboy JR, Garl PJ (1997). HIV-1 LTR DNA sequence variation in brain-derived isolates. *J NeuroVirol* **5:** 331–341.

Cunningham PH, Smith DG, Satchell C, Cooper DA, Brew B (2000). Evidence for independent development of resistance to HIV-1 reverse transcriptase inhibitors in the cerebrospinal fluid. *AIDS* **14:** 1949–1954.

Ding J, Hughes SH, Arnold E (1997). Protein-nucleic acid interactions and DNA conformation in a complex of human immunodeficiency virus type 1 reverse transcriptase with a double-stranded DNA template-primer. *Biopolymers* **44:** 125–138.

Di Stefano M, Gray F, Leitner T, Chiodi F (1996). Analysis of ENV V3 sequences from HIV-1-infected brain indicates restrained virus expression throughout the disease. *J Med Virol* **49:** 41–48.

Ellis RJ, Gamst AC, Capparelli E, Spector SA, Hsia K, Wolfson T, Abramson I, Grant I, McCutchan JA (2000). Cerebrospinal fluid HIV RNA originates from both local CNS and systemic sources. *Neurology* **54:** 927–936.

Epstein LG, Kuiken C, Blumberg BM, Hartman S, Sharer LR, Clement M, Goudsmit J (1991). HIV-1 V3 domain variation in brain and spleen of children with AIDS: Tissue-specific evolution within host-determined quasispecies. *Virology* **180:** 583–590.

Ganeshan S, Dickover RE, Korber BT, Bryson YJ, Wolinsky SM (1997). Human immunodeficiency virus type 1 genetic evolution in children with different rates of development of disease. *J Virol* **71:** 663–677.

Gatanaga H, Oka S, Ida S, Wakabayashi T, Shioda T, Iwamoto A (1999). Active HIV-1 redistribution and replication in the brain with HIV encephalitis. *Arch Virol* **144:** 29–43.

Gisslen M, Hagberg L, Norkrans G, Lekman A, Fredman P (1997). Increased cerebrospinal fluid ganglioside GM1 concentrations indicating neuronal involvement in all stages of HIV-1 infection. *J NeuroVirol* **3:** 148–152.

Glass JD, Wesselingh SL, Selnes OA, McArthur JC (1993). Clinical-neuropathologic correlation in HIV-associated dementia. *Neurology* **43:** 2230–2237.

Guillon C, Bedin F, Fouchier RAM, van't Wout AB, Guters RA (1995). [online] ⟨http://hiv-web.lanl.gov/⟩

Haas DW, Clough LA, Johnson BW, Harris VL, Spearman P, Wilkinson GR, Fletcher CV, Fiscus S, Raffanti S, Donlon R, McKinsey J, Nicotera J, Schmidt D, Shoup RE, Kates RE, Lloyd RM, Jr, Larder B (2000). Evidence of a source of HIV type 1 within the central nervous system by ultraintensive sampling of cerebrospinal fluid and plasma. *AIDS Res Hum Retroviruses* **16:** 1491–1502.

Haas J (1993). [online] ⟨http://hiv-web.lanl.gov/⟩

Halvas EK, Svarovskaia ES, Pathak VK (2000). Role of murine leukemia virus reverse transcriptase deoxyribonucleoside triphosphate-binding site in retroviral replication and *in vivo* fidelity. *J Virol* **74:** 10349–10358.

Huang H, Chopra R, Verdine GL, Harrison SC (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: Implications for drug resistance. *Science* **282:** 1669–1675.

Iftimovici E, Rabian C, Burgard M, Peytavin G, Rouzioux C, Viard JP (1998). Longitudinal comparison of HIV-1 RNA burden in plasma and cerebrospinal fluid in two patients starting triple combination antiretroviral therapy. *AIDS* **12:** 535–537.

Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ (1998). Multiple sequence alignment with Clustal X. *Trends Biochem Sci* **23:** 403–405.

Jonckheere H, De Clercq E, Anne J (2000). Fidelity analysis of HIV-1 reverse transcriptase mutants with an altered amino-acid sequence at residues Leu74, Glu89, Tyr115, Tyr183 and Met184. *Eur J Biochem* **267:** 2658–2665.

Julias JG, Pathak VK (1998). Deoxyribonucleoside triphosphate pool imbalances *in vivo* are associated with an increased retroviral mutation rate. *J Virol* **72:** 7941–7949.

Kaushik N, Talele TT, Pandey PK, Harris D, Yadav PN, Pandey VN (2000). Role of glutamine 151 of human immunodeficiency virus type-1 reverse transcriptase in substrate selection as assessed by site-directed mutagenesis. *Biochemistry* **39:** 2912–2920.

Kim B, Ayran JC, Sagar SG, Adman ET, Fuller SM, Tran NH, Horrigan J (1999). New human immunodeficiency virus, type 1 reverse transcriptase (HIV-1 RT) mutants with increased fidelity of DNA synthesis. Accuracy, template binding, and processivity. *J Biol Chem* **274:** 27666–27673.

Kohlstaedt LA, Wang J, Friedman JM, Rice PA, and Steitz TA (1992). Crystal structures at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256:** 1783–1790.

Korber B (1994a). Computational analysis of HIV molecular sequences. In *HIV signature and sequence variation analysis*. Rodrigo AG, Learn GH (eds). Kluwer Academic Publishers: Dordrecht, Netherlands, pp 55–72.

Korber BT, Kunstman KJ, Patterson BK, Furtado M, McEvilly MM, Levy R, Wolinsky SM (1994b). Genetic differences between blood- and brain-derived viral sequences from human immunodeficiency virus type 1-infected patients: Evidence of conserved elements in the V3 region of the envelope protein of brain-derived sequences. *J Virol* **68:** 7467–7481.

Krebs FC, Mehrens D, Pomeroy S, Goodenow MM, Wigdahl B (1998). Human immunodeficiency virus type 1 long terminal repeat quasispecies differ in basal transcription and nuclear factor recruitment in human glial cells and lymphocytes. *J Biomed Sci* **5:** 31–44.

Lewis DA, Bebenek K, Beard WA, Wilson SH, Kunkel TA (1999). Uniquely altered DNA replication fidelity conferred by an amino acid change in the nucleotide binding pocket of human immunodeficiency virus type 1 reverse transcriptase. *J Biol Chem* **274:** 32924–32930.

Li Y, Kappes JC, Conway JA, Price RW, Shaw GM (1992). [online] ⟨http://hiv-web.lanl.gov/⟩

Liu C, Power C (1996). [online] ⟨http://hiv-web.lanl.gov/⟩

Lukashov VV (2000). [online] ⟨http://hiv-web.lanl.gov/⟩

Mansky LM, Temin HM (1995). Lower *in vivo* mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol* **69:** 5087–5094.

Masliah E, DeTeresa RM, Mallory ME, Hansen LA (2000). Changes in pathological findings at autopsy in AIDS cases for the last 15 years. *AIDS* **14:** 69–74.

Matala E (1996). [online] ⟨http://hiv-web.lanl.gov/⟩

McArthur JC (1987). Neurologic manifestations of AIDS. *Medicine* **66:** 407–437.

McArthur JC, Becker PS, Parisi JE, Trapp B, Selnes OA, Cornblath DR, Balakrishnan J, Griffin JW, Price D (1989). Neuropathological changes in early HIV-1 dementia. *Ann Neurol* **26:** 681–684.

Morris A, Marsden M, Halcrow K, Hughes ES, Brettle RP, Bell JE, Simmonds P (1999). Mosaic structure of the human immunodeficiency virus type 1 genome infecting lymphoid cells and the brain: Evidence for frequent *in vivo* recombination events in the evolution of regional populations. *J Virol* **73:** 8720–8731.

Najera I, Holguin A, Quinones-Mateu ME, Munoz-Fernandez MA, Najera R, Lopez Galindez C, Domingo E (1995). Pol gene quasispecies of human immunodeficiency virus: Mutations associated with drug resistance in virus from patients undergoing no drug therapy. *J Virol* **69:** 23–31.

Nei M, Gojobori T (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3:** 418–426.

O'Brien WA, Koyanagi Y, Namazie A, Zhao JQ, Diagne A, Idler K, Zack JA, Chen ISY (1991). [online] ⟨http://hiv-web. lanl.gov/⟩

Oelrichs RB, Mcphee DA, Deacon NJ (1998). [online] ⟨http://hiv-web.lanl.gov/⟩

Ota T, Nei M (1994). Variance and covariances of the numbers of synonymous and nonsynonymous substitutions per site. *Mol Biol Evol* **11:** 613–619.

Pang S, Vinters HV, Akashi T, O'Brien WA, Chen IS (1991). HIV-1 env sequence variation in brain tissue of patients with AIDS-related neurologic disease. *J Acquir Immune Defic Syndr* **4:** 1082–1092.

Power C, McArthur JC, Johnson RT, Griffin DE, Glass JD, Dewey R, Chesebro B (1995). Distinct HIV-1 env sequences are associated with neurotropism and neurovirulence. *Curr Top Microbiol Immunol* **202:** 89–104.

Sarafianos SG, Das K, Ding J, Boyer PL, Hughes SH, Arnold E (1999). Touching the heart of HIV-1 drug resistance: The fingers close down on the dNTP at the polymerase active site. *Chem Biol* **6:** R137–146.

Shafer RW, Stevenson D, Chan B (1999). Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res* **27:** 348–352.

Shah FS, Curr KA, Hamburgh ME, Parniak M, Mitsuya H, Arnez JG, Prasad VR (2000). Differential influence of nucleoside analog-resistance mutations K65R and L74V on the overall mutation rate and error specificity of human

immunodeficiency virus type 1 reverse transcriptase. *J Biol Chem* **275:** 27037–27044.

Shaw GM, Harper ME, Hahn BH, Epstein LG, Gajdusek DC, Price RW, Navia BA, Petito CK, O'Hara CJ, Groopman JE (1985). HTLV-III infection in brains of children and adults with AIDS encephalopathy. *Science* **227:** 177–182.

Stankoff B, Calvez V, Suarez S, Bossi P, Rosenblum O, Conquy L, Turell E, Dubard T, Coutellier A, Baril L, Bricaire F, Lacomblez L, Lubetzki C (1999). Plasma and cerebrospinal fluid human immunodeficiency virus type-1 (HIV-1) RNA levels in HIV-related cognitive impairment. *Eur J Neurol* **6:** 669–675.

Strizki JM, Albright AV, Sheng H, O'Connor M, Perrin L, Gonzalez-Scarano F (1996). Infection of primary human microglia and monocyte-derived macrophages with human immunodeficiency virus type 1 isolates: Evidence of differential tropism. *J Virol* **70:** 7654–7662.

van't Wout AB, Ran LJ, Kuiken CL, Kootstra NA, Pals ST, Schuitemaker H (1998). Analysis of the temporal relationship between human immunodeficiency virus type 1 quasispecies in sequential blood samples and various organs obtained at autopsy. *J Virol* **72:** 488–496.

Venturi G, Catucci M, Romano L, Corsi P, Leoncini F, Valensin PE, Zazzi M (2000). Antiretroviral resistance mutations in human immunodeficiency virus type 1 reverse transcriptase and protease from paired cerebrospinal fluid and plasma samples. *J Infect Dis* **181:** 740–745.

Weiss KK, Isaacs SJ, Tran NH, Adman ET, Kim B (2000). Molecular architecture of the mutagenic active site of human immunodeficiency virus type 1 reverse transcriptase: Roles of the beta 8-alpha E loop in fidelity, processivity, and substrate interactions. *Biochemistry* **39:** 10684–10694.

Wildemann B, Haas J, Ehrhart K, Wagner H, Lynen N, Storch-Hagenlocher B (1993). *In vivo* comparison of zidovudine resistance mutations in blood and CSF of HIV-1-infected patients. *Neurology* **43:** 2659–2663.

Wong JK, Ignacio CC, Torriani F, Havlir D, Fitch NJ, Richman DD (1997). *In vivo* compartmentalization of human immunodeficiency virus: Evidence from the examination of pol sequences from autopsy tissues. *J Virol* **71:** 2059–2071.

Wrobel JA, Chao SF, Conrad MJ, Merker JD, Swanstrom R, Pielak GJ, Hutchison CA (1998). A genetic approach for identifying critical residues in the fingers and palm subdomains of HIV-1 reverse transcriptase. *Proc Natl Acad Sci USA* **95:** 638–645.