AIDS Reviews 2004;6:54-60

Lost in Translation: Implications of HIV-1 Codon Usage for Immune Escape and Drug Resistance

Gustavo H. Kijak[1], Jeffrey R. Currier[1], Sodsai Tovanabutra[1], Josephine H. Cox[1], Nelson L. Michael[2], Scott A. Wegner[2], Deborah L. Birx[2], and Francine E. McCutchan[1]

[1]Henry M. Jackson Foundation for the Advancement of Military Medicine, Rockville, Maryland; [2]Division of Retrovirology, Walter Reed Army Institute of Research, Rockville, Maryland

Abstract

Synonymous nucleotide substitutions in protein-coding sequences are often regarded as evolutionarily neutral and not subject to selective pressure. However, synonymous codons can sometimes lead to different patterns of amino acid substitution by single nucleotide changes. Based on the deconstruction of the standard genetic code, we propose the term 'quasi-synonymous' to describe codons that specify the same amino acid, but lie on different mutational pathways, and we show that in at least one rapidly evolving organism, HIV-1, quasi-synonymy plays a role in its evolution. We present concrete examples that demonstrate the relevance of codon usage in the development of antiretroviral-drug resistance. In the case of the host immune response, the data indicates that viral evasion is achieved through use of codons that lie on the direct path to escape mutants, and equally, permit rapid reversion to wild-type in the absence of these selective pressures. Quasi-synonymy conditions HIV-1 and, potentially, other rapidly evolving organisms in their exploration of the mutational space. **(AIDS Rev 2004;6:54-60)**

Key words

HIV evolution. Condon usage. Immune escape. Drug resistance.

Introduction

This year marks the 20th anniversary of the identification of HIV as the etiological agent of AIDS[1-4]. In two decades we have witnessed the extensive evolution of an organism and of our knowledge about it[5]. HIV-1 has an extreme phenotypic plasticity. It can productively infect different cellular compartments within the host[6], escape pressure from humoral and cellular immune responses[7-9], and replicate in the presence of antiretroviral drugs through development of resistant strains[10-12]. This is achieved through a high mutation rate[13], recombination[14], and rapid replication within the host[15,16]. Like some other RNA viruses, HIV-1 generates a quasispecies, a swarm of related but distinct viruses, through which it explores the mutational space and occupies a large adaptive landscape[17,18].

HIV-1 quasispecies generate all possible single-nucleotide mutations, and even some multiple mutants, in a finite time[19]. Concepts in molecular genetics developed from organisms with lower evolutionary rates may not always apply to HIV-1. Here we discuss the implications of rapid mutation for one of the most fundamental of these: codon synonymy in the genetic code.

Degeneracy in the genetic code produces disparate results for single nucleotide substitutions; some lead to a change in the encoded amino acid (non-synonymous) while others substitute a different codon for the same amino acid (synonymous)[20,21]. In protein-coding genes, positions under positive selection can be identified by a higher rate of non-synonymous substitutions, often expressed as the ratio dn/ds[22,23]. This concept has been

Correspondence to:
Gustavo H. Kijak
Henry M. Jackson Foundation,
1600 East Gude Drive
Rockville, MD 20850
Phone: 1-301-251-5046
Fax: 1-301-762-7460
E-mail: gkijak@hivresearch.org

applied to HIV-1 to identify and map selection by the host immune response and by antiretroviral-drugs[24-26].

The term synonymous reflects the perception that alternative codons for an amino acid are equivalent in the evolutionary sense, that is, substitution of one for another has no direct consequence on phenotype. However, there is strong evidence that codons themselves are also under selection. Bias in codon usage is often observed, even among synonymous codons[27]. Modulation of translation efficiency due to the relative abundance of iso-accepting aminoacyl t-RNAs[28], G + C levels conditioned by gene location in chromosomes[29], bias towards overall nucleotide composition[30], and suppression of the CpG dinucleotides in eukaryotic viruses[31], have been among the processes invoked to account for bias in codon usage.

The key fact about a rapidly evolving organism like HIV-1 is that mutations can occur in series within a single codon. Thus the individual codon represents a mutational pathway that HIV-1 can explore. The fact that synonymous codons for some amino acids lie on different mutational pathways, usually of no consequence because mutation rates and generation times in most organisms are so slow that the likelihood of two mutations in the same codon is insignificant, comes into play for HIV-1. As figure 1 illustrates, among the four glycine codons, GGG is the only one that can change to the tryptophan codon TGG by a single mutation. At every position where glycine occurs, utilization of GGG to encode it provides a pathway to tryptophan that the other three codons do not. The codon pair GGT and GGC can mutate to codons for aspartic acid, cysteine or serine by a single mutation, while the pair GGA and GGG cannot, yet all encode glycine; the two pairs of codons lie on different mutational pathways. Even more subtle forms of asymmetry can be observed. GGA and GGG have two pathways to arginine (GGA/GGG to CGA/CGG, and GGA/GGG to AGA/AGG, see thick arrow, Fig. 1), while GGC and GGT have only one; substitution of arginine for glycine would, all other factors being equal, be more frequent using the first pair of codons, yet for an alanine or valine substitution, all glycine codons provide equivalent access. On the other hand, as the mutational process is reversible, one can assume that, if a glycine had derived from a tryptophan, the codon usage for that glycine would most likely be GGG, while if it had derived from a serine, the glycine-coding codons would be GGC or GGT; back mutations access the mutational pathway used for forward mutation. The asymmetries observed for glycine do not pertain to all amino acids. Figure 1 shows that all point mutations in the two histidine codons lead to the same spectrum of amino acid substitutions.

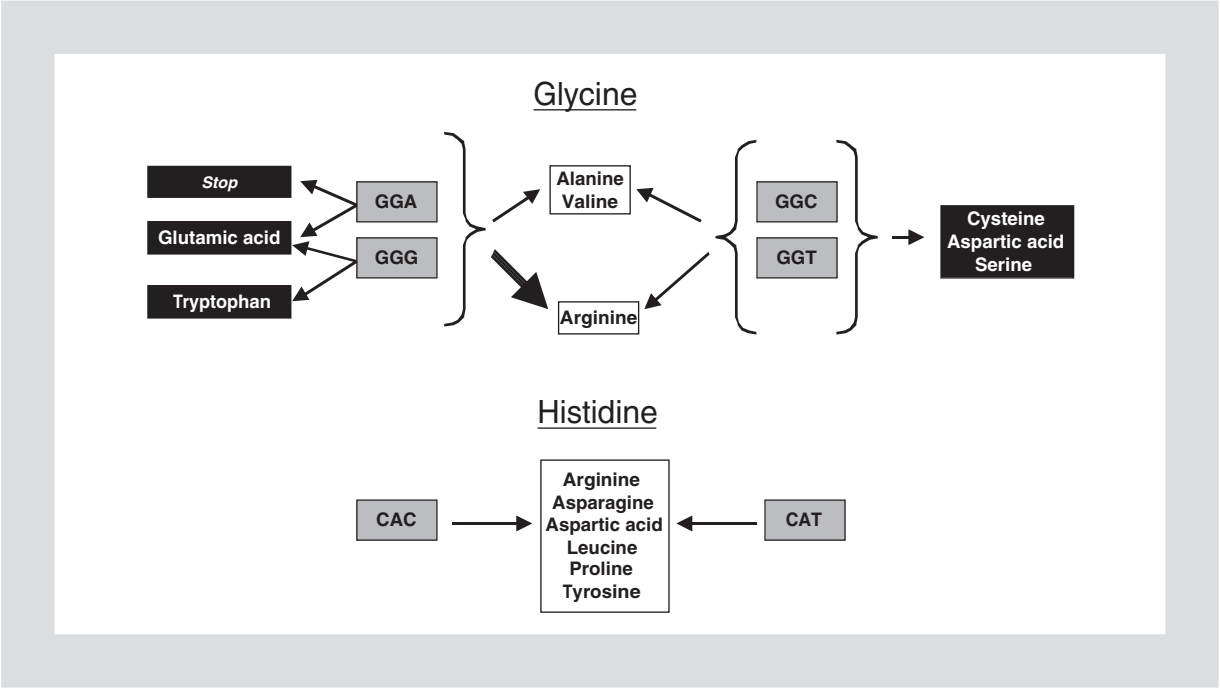

Figure 1. *Synonymous codons can be involved in different mutational pathways. Amino-acidic changes resulting from single nucleotide substitutions in glycine and histidine specifying codons (in gray) are shown. Convergent and divergent destinations are indicated in white and black boxes, respectively. The thick arrow represents two different pathways through which GGA and GGG can mutate into arginine.*

The terms synonymous and non-synonymous do not capture these asymmetries in mutational pathways. We propose the term **'quasi-synonymy'** to describe the relationship between codons that specify the same amino acid but can achieve a different pattern of amino acid substitution by single nucleotide changes. In the examples shown in figure 1, the glycine codons GGC and GGT are synonymous, but all other pairs of glycine codons show only partial synonymy; they are quasi-synonymous. Histidine provides an example of codons that are fully synonymous

Figure 2 presents a deconstruction of the standard genetic code according to the degree of quasi-synonymy. At one extreme lie amino acids for which single mutations in all of their codons lead to the same spectrum of amino acid substitutions; the alternative codons are truly synonymous for phenylalanine, histidine, glutamine, glutamic acid, cysteine, aspartic acid, asparagine, and tyrosine. On the other extreme we can find amino acids for which each codon leads to a different spectrum of amino acid substitutions; arginine, lysine and serine are in this group. For alanine, lysine and proline, the codons are almost synonymous, as their mutational pathways share most of the possible amino acid substitutions, whereas threonine, glycine, isoleucine, and valine codons are mostly quasi-synonymous; each lies on a different mutational pathway.

The fact of quasi-synonymy in the genetic code seems incontrovertible. The question before us is this: does it have relevance for the evolution of HIV-1 and, by analogy, for other rapidly evolving organisms? We turn to two of the best-studied selective pressures on HIV-1 – antiretroviral drugs and the immune response – to provide concrete examples of its importance.

Quasi-synonymous codons and the development of drug resistance

Efavirenz (EFV) is a non-nucleoside reverse-transcriptase inhibitor (NNRTI) for treatment of HIV-1 infection[32]. It has been used against HIV-1 subtype B, the main circulating strain in North America, Europe, and Australia[33,34], and subtype C, the predominant strain in southern Africa[35]. A valine at position 106 of reverse transcriptase (RT) is mainly encoded by GTA in subtype B, but by GTG in subtype C[36]. These valine codons are considered synonymous but, in fact, they are on different mutational pathways (Fig. 2); one leads to methionine and the other to isoleucine. This 'silent polymorphism' proved crucial to the treatment outcome: individuals infected with the C subtype generated V106M, resistant not only to EFV but to all members of the NNRTI family, while mutations at V106 were rarely observed in infections with subtype B[37].

We evaluated the database of HIV-1 sequences for other examples of the potential impact of quasi-synonymy. Mutation L210W in RT contributes to high-level zidovudine (AZT) resistance in the presence of mutations T215Y and M41L[38,39]. Among six leucine codons, only TTG lies on the direct mutational pathway to tryptophan. Among 1306 subtype B sequences, leucine 210 is mainly encoded by TTG, while in 334 subtype C strains, about half have TTG and the others have TTA[40]. The effect of this could be development of a lower level of AZT resistance in some subtype C strains, or selection for a different pathway to high-level resistance.

The important mutation Q151M emerges in up to 16% of patients treated with AZT in combination with either zalcitabine (ddC) or didanosine (ddI), conferring cross-resistance to the whole class of nucleoside RT inhibitors (NRTIs)[41]. The pathway involves leucine as an intermediate (Q to L to M)[42]. Q151 is encoded by CAG in 91% of 942 subtype B strains, providing the pathway Q_{CAG} to L_{CTG} to M_{ATG}[40]. In subtype D in East Africa, among 181 strains only 47% have CAG this position; 53% have CAA instead, for which the path to methionine through leucine is longer (Q_{CAA} to L_{CTA} to L_{CTG} to M_{ATG}). In subtype D infections, resistance may develop more slowly or take a different path.

The functional importance of quasi-synonymy has been revealed by the spotlight of drug pressure. Resistance mutations not detected in subtype B could be just one mutation away from wild-type in other subtypes. The more global use of antiretrovirals may call forth the hidden potential in quasi-synonymy.

Quasi-synonymous codons and the escape from the immune response

Strong selective pressures against HIV-1 are generated by cellular immune responses, particularly cytotoxic T-lymphocytes (CTL)[9,43]. CTL responses are restricted by the highly polymorphic class I MHC complex in human populations; during the course of infection, the spectrum of escape mutants generated by HIV-1 is focused on those epitopes that are recognized by the specific HLA type of the host[8]. At the time of transmission, HIV-1 usually faces a new set of pressures brought to bear by a new HLA type; many of the old 'escape mutants' are no longer useful and new ones need to be quickly generated in order to

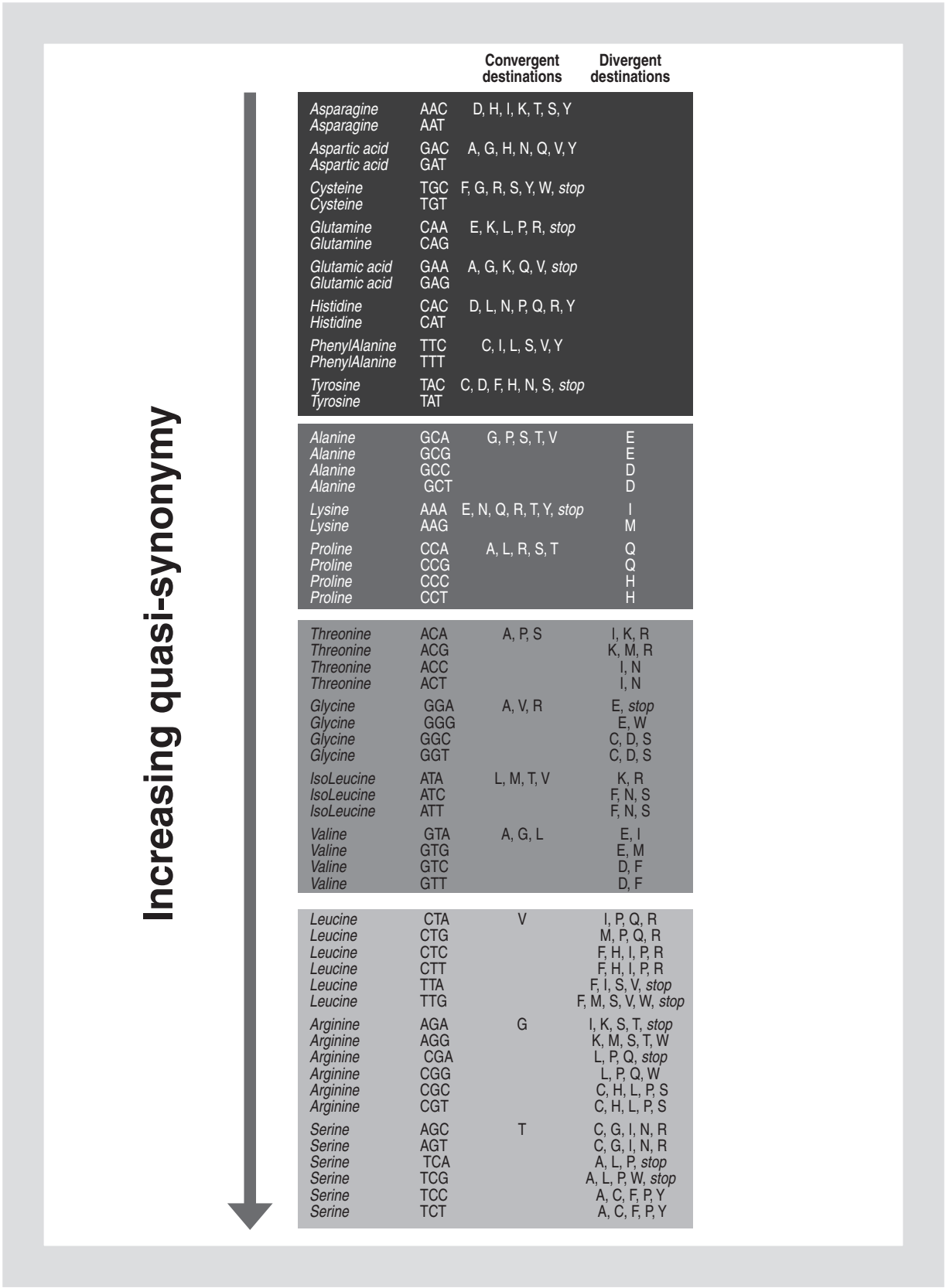

Figure 2. *Deconstruction of the standard genetic code based on the degree of quasi-synonymy of codons specifying for the same amino acid. Convergent and divergent destinations achievable through single nucleotide substitutions are shown for the codons of the same amino-acid coding family. Methionine and tryptophan are not shown as only one codon in each case specifies for these residues.*

survive. At the same time, escape mutants may come with a penalty in terms of viral fitness; survival in polymorphic human populations must depend to some degree on the plasticity to alternate between phenotypes.

Could codon selection contribute to this plasticity? In light of quasi-synonymy, one simple solution for HIV-1 would be to select codons on the mutational pathways connecting escape mutants to wild-types. We examined the HIV-1 sequence database for subtypes B and C[44] with respect to a well-known CTL epitope, TLYCVHQR in position 84-91 of Gag presented by the A*01101 class I HLA allele[45,46]. The C-terminal anchor of this epitope is at position 91 where arginine or lysine is tolerated[47]. The most frequent amino acids at this position in subtype B are arginine and lysine (Fig. 3) but, in subtype C, the most common residue is glycine, almost certainly an escape mutant because small, polar or negatively charged amino acids are uncommon in C-terminal anchor positions[48]. Among the six codons for arginine and the four for glycine, there is almost complete dominance of those that lie on the pathway connecting arginine, lysine and glycine by single nucleotide substitutions. Arginine and lysine, on the one hand, and glycine on the other, may represent different phenotypic states, in this case wild-type presented by A*01101 and escape mutants, whose plasticity is maintained through codon selection. Host genetic backgrounds, expressed in different HLA types and allele frequencies, may have carved the observed pathway into the genome of HIV-1, exploiting quasi-synonymy to stay on the reversible pathway between wild-type and CTL[49] escape mutants.

Conclusion

Most studies of evolution consider the amino acid to be the unit of selection, responsible for phenotypic

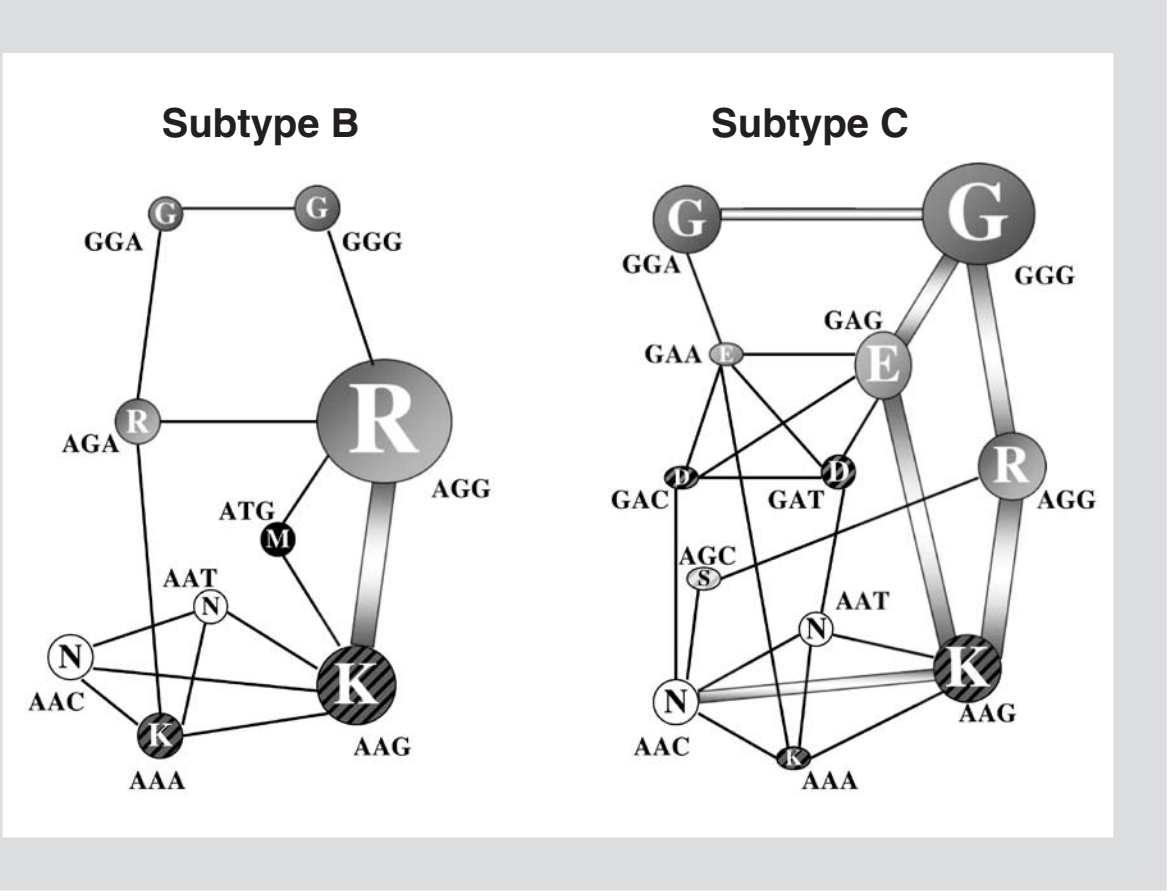

Figure 3. *HIV-1 mutational plasticity and escape from immune response achieved through the usage of quasi-synonyms. Mutational pathways linking the different codons that specify for the residues present in subtype B (n = 176) and subtype C (n = 149) HIV-1 isolates at position 91 of Gag[44] are shown. This position constitutes the C-anchor of the epitope TLYCVHQR, presented by the A*01101 class I HLA allele[45,46]. The relative size of the circles represents the frequency of the codon in each subtype. The thick lines link the most frequent polymorphisms. See text for details.*

change. Thus, information contained in the genotype has been partially discounted, sweeping aside of synonymous codons as 'silent' mutations. This bias hid from our view a dynamic process: the utilization of quasi-synonymy to establish mutational pathways, in rapidly evolving organism. This is a third dimension of mutational space, and to understand it fully we must capture this new dynamic hidden in the genetic code.

In rapidly evolving organisms like HIV-1, past, present, and future are simultaneously present when viewed through the eyes of slowly evolving organisms like us. The polymorphisms that we observe at positions under strong selective pressure are but a still image of a rapidly moving object. The current state of an HIV-1 strain is linked to its past and future through the conduit of quasi-synonymy, fixing some codons that are on the pathway to reversible change. The consequence of this fact is that at least some of the evolution of HIV-1 is predictable from its codon usage. Could this be exploited to produce more effective vaccines or superior antiretroviral drugs? The plasticity of HIV-1 is often invoked as a major barrier to its control[50], yet the plasticity of the virus may be more limited than heretofore appreciated. In exploring the mutational space, HIV-1 has been molded, not into an optimal strain, but into a dynamic machine, poised to move in certain directions and equally prevented from moving in others.

References

1. Barre-Sinoussi F, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 1983;220:868-71.
2. Gallo R, et al. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). Science 1983;220:865-7.
3. Gallo R. Historical essay. The early years of HIV/AIDS. Science 2002;298:1728-30.
4. Montagnier L. Historical essay. A history of HIV discovery. Science 2002;298:1727-8.
5. Fauci A. HIV and AIDS: 20 years of science. Nat Med 2003;9: 839-43.
6. Berger E, Murphy P, Farber J. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. Annu Rev Immunol 1999;17:657-700.
7. Mascola J, Montefiori D. HIV-1: nature's master of disguise. Nat Med 2003;9:393-4.
8. Klenerman P, Wu Y, Phillips R. HIV: current opinion in escapology. Curr Opin Microbiol 2002;5:408-13.
9. Brander C, Walker B. Gradual adaptation of HIV to human host populations: good or bad news? Nat Med 2003;9:1359-62.
10. Deeks S. International perspectives on antiretroviral resistance. Nonnucleoside reverse transcriptase inhibitor resistance. J Acquir Immune Defic Syndr 2001;26(Suppl 1):25-33.
11. Miller V. International perspectives on antiretroviral resistance. Resistance to protease inhibitors. J Acquir Immune Defic Syndr 2001;26(Suppl 1):34-50.
12. Loveday C. International perspectives on antiretroviral resistance. Nucleoside reverse transcriptase inhibitor resistance. J Acquir Immune Defic Syndr 2001;26(Suppl 1):10-24.
13. Mansky L, Temin H. Lower *in vivo* mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J Virol 1995;69:5087-94.
14. Burke D. Recombination in HIV: an important viral evolutionary strategy. Emerg Infect Dis 1997;3:253-9.
15. Wei X, et al. Viral dynamics in human immunodeficiency virus type 1 infection. Nature 1995;373:117-22.
16. Ho D, et al. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature 1995;373:123-6.
17. Eigen M, Biebricher C. Sequence space and quasispecies distribution. In: Domingo E, Holland J, Ahlquist P (eds). RNA genetics. Vol III. Boca Raton, Fla: CRC Press Inc1998:3-22.
18. Holland J, De La Torre J, Steinhauer D. RNA virus populations as quasispecies. Curr Top Microbiol Immunol 1992;176:1-20.
19. Coffin J. HIV population dynamics *in vivo:* implications for genetic variation, pathogenesis, and therapy. Science 1995;267:483-9.
20. Nei M, Kumar S. Molecular evolution and phylogenetics. Oxford, UK: Oxford University Press 2000.
21. Lewin B. Genes VIII. Oxford, UK: Oxford University Press 2000.
22. Yang Z, Bielawski J. Statistical methods for detecting molecular adaptation. Trends in Ecology and Evolution 2000;15:496-503.
23. Hughes A. Adaptative evolution of genes and genomes. Oxford, UK: Oxford University Press 1999.
24. Soudeyns H, et al. Selective pressure exerted by immunodominant HIV-1-specific cytotoxic T lymphocyte responses during primary infection drives genetic variation restricted to the cognate epitope. Eur J Immunol 1999;29:3629-35.
25. Yang W, Bielawski J, Yang Z. Widespread adaptive evolution in the human immunodeficiency virus type 1 genome. J Mol Evol 2003;57: 212-21.
26. Sa-Filho D, et al. Analysis of the protease sequences of HIV-1 infected individuals after Indinavir monotherapy. J Clin Virol 2003;28: 186-202.
27. Grantham R, Gautier C, Gouy M, Mercier R, Pave A. Codon catalog usage and the genome hypothesis. Nucleic Acids Res 1980;8: r49-r62.
28. Ikemura T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J Mol Biol 1981;151:389-409.
29. Bernardi G. The vertebrate genome: isochores and evolution. Mol Biol Evol 1993;10:186-204.
30. Van Hemert F, Berkhout B. The tendency of lentiviral open reading frames to become A-rich: constraints imposed by viral genome organization and cellular tRNA availability. J Mol Evol 1995;41:132-40.
31. Oliver J, Marin A. A relationship between GC content and coding-sequence length. J Mol Evol 1996;43:216-23.
32. Adkins J, Noble S. Efavirenz. Drugs 1998;56:1055-64; discussion 1065-6.
33. McCutchan F. Understanding the genetic diversity of HIV-1. Aids 2000;14(Suppl 3):31-44.
34. Peeters M, Sharp PM. Genetic diversity of HIV-1: the moving target. Aids 2000;14(Suppl 3):129-40.
35. Van Harmelen J, et al. A predominantly HIV type 1 subtype C-restricted epidemic in South African urban populations. AIDS Res Hum Retroviruses 1999;15:395-8.
36. Loemba H, et al. Genetic divergence of human immunodeficiency virus type 1 Ethiopian clade C reverse transcriptase (RT) and rapid development of resistance against nonnucleoside inhibitors of RT. Antimicrob Agents Chemother 2002;46:208794.
37. Brenner B, et al. A V106M mutation in HIV-1 clade C viruses exposed to efavirenz confers cross-resistance to non-nucleoside reverse transcriptase inhibitors. Aids 2003;17:1-5.
38. Gurusinghe A, et al. Reverse transcriptase mutations in sequential HIV-1 isolates in a patient with AIDS. J Med Virol 1995;46:238-43.
39. Hooker D, et al. An *in vivo* mutation from leucine to tryptophan at position 210 in human immunodeficiency virus type 1 reverse tran-

scriptase contributes to high-level resistance to 3'-azido-3'-deoxythymidine. J Virol 1996;70:8010-8.
40. Rhee S, et al. Human immunodeficiency virus reverse transcriptase and protease sequence database. Nucleic Acids Res 2003;31: 298-303.
41. Shirasaka T, et al. Emergence of human immunodeficiency virus type 1 variants with resistance to multiple dideoxynucleosides in patients receiving therapy with dideoxynucleosides. Proc Natl Acad Sci USA 1995;92:2398-402.
42. García-Lerma J, Gerrish P, Wright A, Qari S, Heneine W. Evidence of a role for the Q151L mutation and the viral background in development of multiple dideoxynucleoside-resistant human immunodeficiency virus type 1. J Virol 2000;74:9339-46.
43. McMichael A, Rowland-Jones S. Cellular immune responses to HIV. Nature 2001;410:980-7.
44. HIV Sequence Database LANL, Los Alamos, NM (http://hiv-web.lanl.gov). Accessed 12 November 2003.
45. Harrer T, et al. Recognition of two overlapping CTL epitopes in HIV-1 p17 by CTL from a long-term non-progressing HIV-1-infected individual. J Immunol 1998;161:4875-81.
46. Fukada K, et al. Cytotoxic T-cell recognition of HIV-1 cross-clade and clade-specific epitopes in HIV-1-infected Thai and Japanese patients. Aids 2002;16:701-11.
47. Kubo R, et al. Definition of specific peptide motifs for four major HLA-A alleles. J Immunol 1994;152:3913-24.
48. Yusim K, et al. Clustering patterns of cytotoxic T-lymphocyte epitopes in human immunodeficiency virus type 1 (HIV-1) proteins reveal imprints of immune evasion on HIV-1 global variation. J Virol 2002;76:8757-68.
49. Moore C, et al. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. Science 2002;296:1439-43.
50. Thomson M, Pérez-Álvarez L, Nájera R. Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. Lancet Infect Dis 2002;2:461-71.