Modeling recombination's role in the evolution of HIV drug resistance *

Richard K. Belew, Max W. Chang
Cognitive Science Dept., Bioinformatics Program
Univ. California – San Diego

12 January 2006

This is a slightly expanded version of the paper to appear in Artificial Life X, June 3-7, 2006 Bloomington, IN USA. See `www.alifex.org`*. It differs from the refereed, published version only in expanded layout, color versions of figures and several citations that were elided because of length restrictions.*

Abstract

HIV is obviously one of the most deadly diseases currently facing our species, but it also provides an extraordinary, well-documented example of a "successful" evolutionary system. We present a simple model that is capable of exploring interactions among several key features of this system that push the limits of traditional evolutionary theory. First, HIV turns out to exploit recombination, a phenomenon typically associated only with sexually reproducing species and also a controversial aspect of many evolutionary simulations. Second, it exhibits "phenotypic mixing," allowing genotypic information to be "hidden" within phenotypic virions (viral particles). Third, current drug therapies designed to attack the virus typically include *combinations* of multiple drugs, to which the virus has successfully evolved resistant strains in many cases. Finally, recent analyses of "effective" population sizes of the virus make it much smaller than "census" populations might lead one to believe, and also making simulations of stochastic effects worthwhile. This paper describes basic features of the model and reports several preliminary results, including new evidence of an interaction between recombination and phenotypic mixing.

1 Introduction

As with many other biological phenomena, the human immunodeficiency virus HIV brings together an extraordinarily complex combination of inter-related

*`http://www.cogsci.ucsd.edu/~rik/papers/hivPop-alifeX.pdf`

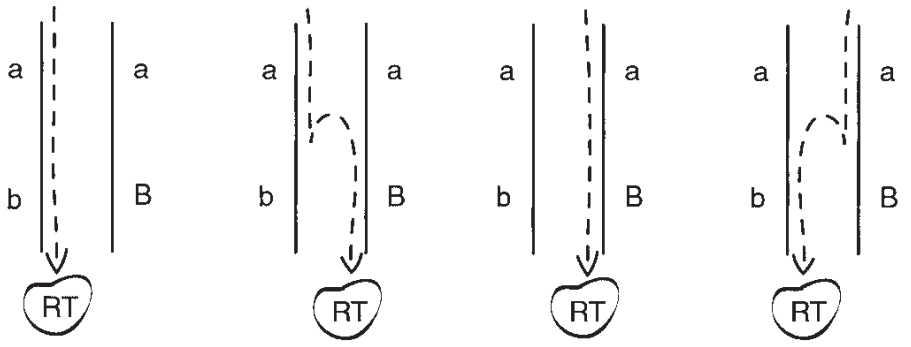


Figure 1: Retroviral recombination, from Bretscher et al.

dynamics. Viral infection of macrophages (e.g., brain microglia) does not cause these cells to die, and therefore these become a reservoir for virus production. Within these "refugia," HIV is capable of evolving at a prodigious rate, producing on the order of 10^{10} virions daily. Worse, the reverse transcription process exploited by HIV as it introduces its DNA into host cells creates extremely high mutation rates, affecting as many as one of every 2000 nucleotides. Once infected by a wild-type variant of the disease, drug-resistant mutants can be transmitted to newly infected individuals. Currently 14,000 new infections occur daily, with 95% of these occurring in the developing world least able to afford the drugs that have been developed. The importance of controlling the scourge of HIV is motivation enough. But this remarkable evolving system has at least four other characteristics that make it important for students of evolutionary computation generally.

Recombination

It turns out that HIV exploits a form of recombination similar to that typically associated only with the genetics of sexual species:

> HIV, like all retroviruses, is essentially diploid, since every virion contains two copies of the complete RNA genome. Following the infection of a cell, the reverse transcriptase (RT) attaches to one RNA strand and transcribes the genomic RNA into proviral DNA. During this process, the RT carrying the nascent DNA provirus can fall off its current RNA template and reattach to the other. [Bretscher et al., 2004]

Bretscher et al. also provide a sketch of this process, shown in Figure 1.

The role of a recombination or "crossover" operator as a component of evolutionary search has long been a topic of central concern. Holland [Holland, 1975] and others have highlighted its ability to explore, via "implicitly parallel" search,

vast combinatorial spaces much more effectively than mutation-only alternatives. Others have pointed to the strong constraints on genomic representations required to ensure the formation of "building blocks" supporting "recombinatorial" search. While there have been many empirical demonstrations of the effectiveness of crossover, it is also true that the matter remains controversial.

It is therefore interesting to note that the question of the role of recombination in the evolution of cross-resistant mutants in HIV is also currently widely debated. Previous experimental work has shown that recombination can accelerate the development of HIV drug resistance under specific conditions [Moutouh et al., 1996, Kellam and Larder, 1995]; theoretical advantages of recombination (e.g., escape from Muller's Ratchet, Evolutionary Broad Jumping) have been noted as well [Burke, 1997]. More recently, however, the HIV population modeling by Bretscher et al. has called into question the evolutionary advantage of recombination. A related study of HIV sequences and associated fitness values provided strong evidence for positive epistasis in HIV-1, a characteristic inconsistent with the prevailing view [Bonhoeffer et al., 2004].

Phenotypic mixing

Another dimension of interest to evolutionary algorithm developers concerns the relationship between genetic and phenotypic representations [Banzhaf, 1994, Altenberg, 1995]. Genetic representations – those operated on by genetic operators – can be distinguished from phenotypic representations that are actually evaluated with respect to fitness; there have been many examples of "developmental" functions explicitly transforming genotypic representations to phenotypic ones [Hart et al., 1994]. included grammars [Kammeyer et al., 1995], cellular encoding [Gruau and Whtitley, 1993] and genetic programming [Koza, 1992].

Sidney Brenner first identified a phenomenon in which bacterial phage can exploit this distinction to "hide" genetic information that is distinct from its phenotype [Brenner, 1957]. *Phenotypic mixing* refers to the random packaging of RNA and proteins derived from multiple proviruses within the same cell such that a virion may have fitness associated with a phenotype different than the genotype it also contains. This process is sketched in Figure 2. A set of one or more virions (shown along the top) may infect a cell. New virions are packaged from across this set, and released (bottom). An analysis of the selective dis/advantages associated with distinct genotypic and phenotypic representations arise naturally in the setting of phenotypic mixing.

Co-evolution with respect to drug combination therapies

HIV has been the target of broad-based drug design efforts for almost two decades, and a large number of drug treatments have demonstrated clinical success. These drugs fall in several classes based on the stage of the HIV life cycle they attack, including nucleoside and non-nucleoside classes of reverse transcriptase inhibitors (NRTI, NNRTI), protease inhibitors (PrI), and more recently entry (fusion) inhibitors.

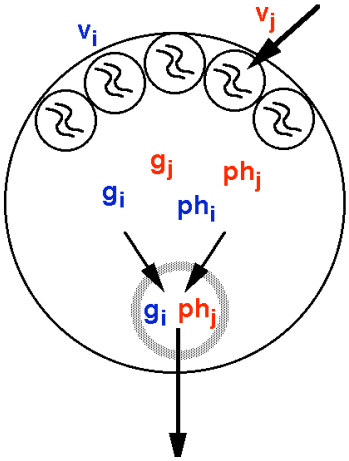


Figure 2: Phenotypic mixing

Based on prior work, our group is most familiar with protease inhibitors. These are especially closely related to structural characterization, primarily due to the "active site" structural characteristics in. NNRTI have many of the same structural, "lock and key" features. NNRTI work by a much different, dynamic mechanism that is less well understood and much more difficult to model. Finally, envelope solutions have a much more recent introduction, and therefore much less available data on resistance.

Highly active anti-retroviral therapy (HAART) refers to multi-drug "cocktail" treatments that use combinations of various inhibitors. These have proven especially effective in blocking viral replication in many patients, at least temporarily. However, for reasons ranging from poor bio-availability to intermittent compliance by infected patients, many patients experience a recurrence of active infection within a few years of first receiving HAART treatment. Resistance to many of the drugs, across all the major drug classes, is therefore also rapidly developing. Evolution of "super mutants" resistant to multiple drugs, and perhaps to all known inhibitors, seems almost inevitable.

Analysis of the co-evolutionary dynamic between medical science's arsenal of potential therapies and the world-wide evolution of the disease in response to these therapies provides an important example for the exploration of drug resistance strategies more generally.

Constructive role for simulation

Finally, the role of computer simulation in the development of theoretic models of disease evolution within individuals and across the human population seems especially promising. Since at least the time of Fisher and Wright, population genetic models have generally depended on mathematical models in which either population sizes are assumed to be sufficiently large that changes

in allele frequencies are "deterministic," or on ones in which smaller "effective" population sizes are imagined to be "stochastic" and described by Kolmogorov's forward equation [Ewens, 1979]. More recently Kimura's theory of "neutral drift," characterization of "neutral landscapes," [Schuster, 1997, MA et al., 1996] etc. have highlighted the role random genetic drift can play. Rouzine et al. [Rouzine et al., 2001] have identified an intermediate regime where selective pressures (s), mutation rates (μ) and effective populations sizes (N_e) are balanced to produce "selection-drift":

$$1/s < N_e < 1/\mu$$

While estimates of the HIV "census" populations are vast, with daily generation of 10^{10} virions, a number of features suggest effective populations closer to 10^3 [Leigh Brown and Richman, 1997], a regime in which stochastic evolution is much more likely. Simulations studies at this critical threshold between deterministic and stochastic regimes are therefore particularly worthwhile. That effective population sizes are in this range also makes simulation of the evolution of reasonable populations, over realistic time periods, computationally tractable.

The next section describes the model, and results of preliminary experiments are presented. The paper concludes with reference to some related work.

2 General model

The two central components of this model are *cells* and *virion particles*. Cells of various types can be simulated as can spatial/topological properties of the organizations of cells within various tissue types. Intra-cell migration of virus within cells of the same tissue type are modeled as flowing with respect to the cells spatial organization. Cross-tissue migrations are mediated by a shared serum providing uniform exchange among all cell types. This is shown in Figure 3. In the experiments presented here, only a single tissue type is considered. Localizing conditions, relative to distribution via serum can expected to allow much more genetic variation due to significantly reduced effective population sizes.

As mentioned above and sketched in Figure 2, phenotypic mixing is also modeled. New virions are produced as proteins associated with genes are each expressed independently, and then RNA and proteins are randomly packaged. In the experiments presented here, a "null" developmental function is used: phenotypes are identical to their genotypes. Note that the random selection of genotype and phenotype still allows (and in fact provides an especially straightforward model of) phenotypic mixing. Statistics are maintained on the number of "homogeneous" virions in which the proteins produced are consistent with their genetic complement and "heterogeneous" virions in which they differ. Because these are simulations, precise counts can also be maintained of those selective events in which differences in genotypic vs. phenotypic fitness "matters": i.e., how often the virion's genotype survives, when it would have been

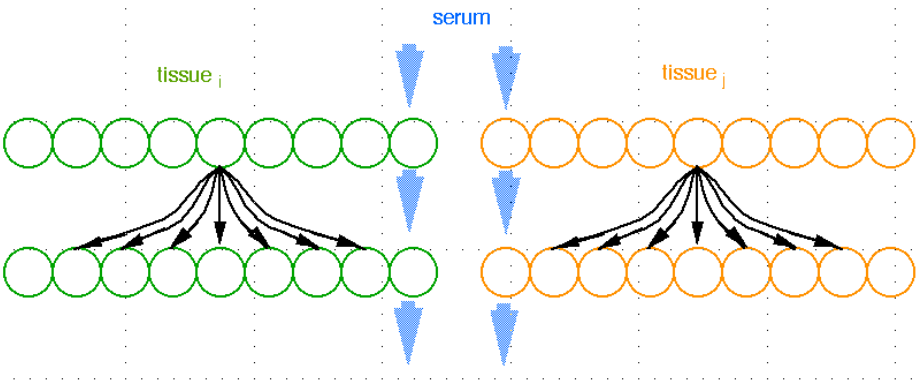


Figure 3: Viral dissemination via serum and tissue

killed if the phenotype actually corresponding its genotype had been evaluated rather than the phenotypic complement behind which it is masked ($Ph > G$), or the converse ($G > Ph$).

2.1 Genomic arrangement

A key feature of recombination is that the probability that two genes become segregated increases with the distance between them. For example, both the protease and reverse transcriptase enzymes are expressed by the same gene (Pol), which is at significant distance from the envelope (Env) gene targeted by fusion inhibitors.

Critical points in the progression of the disease, first prior to any drug treatment, then response to the drug, evolution of resistance to the drug, and final recovery of additional fitness with compensatory mutations, are shown in Figure 4. Our model assumes that a mutation subsequent to the first resistance-producing mutation can restore the virus to some fraction of original, wild-type fitness. Covariational analysis of mutations often suggest relationships between (1x) mutations at one position which seem to have a second, corresponding (2x) "compensatory" mutation at a second position returning the fitness of the virus to new near-wild-type levels [Shafer and Schapiro, 2005]. Multiple databases are beginning to collect both genotypic and phenotypic assays of the fitness of known mutants, and in some cases drug-specific variations as well. Our long-term project is directed towards exploiting exactly this type of empirical data.

For now we consider an especially simplified model with a binary genome and two bits associated with each drug to be considered. These bits are interpreted as shown below:

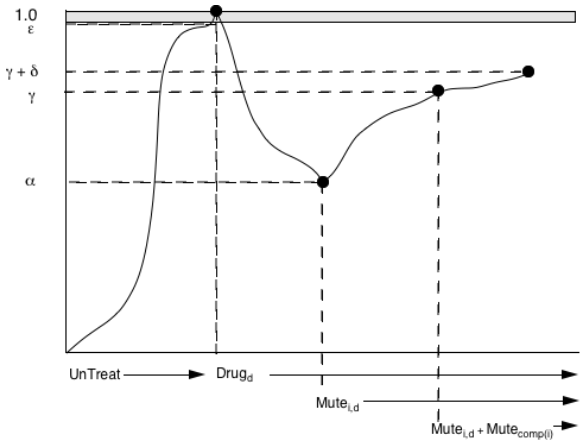


Figure 4: Viral fitness as effected by drug treatment

Bits	Interpretation	Naive Fit.	Drug Fit.
00	wild-type	1.0	0.3
01	wild-type	1.0	0.3
10	Drug resistant	0.95	0.6
11	Drug resist, compen	0.95	0.9

Placement of these genes is also simplified to distinguish only between three, hierarchic distances among three key genetic features of special interest, shown in Figure 5: sequence associated with each subsystem is more proximate on the genome than across subsystems, and the locus of the compensatory mutation (denoted by an overbar in the figure) associated with each resistance mutation are closer still.

In the experiments presented here, we further assume each viral sub-system can be subject to two potential drug therapies, with each assumed to be from a different drug class. Both are modeled as 70% effective against wild-type. Resistance mutation are assumed to double fitness and compensatory mutation are assumed to return the double mutants to 90% effectiveness. Interactions among drug specific fitnesses is modeled as a simple *product* of the individual drugs' fitnesses.

3 Results

In the experiments presented here, an eight-bit genome is used, representing two bits for each of four drugs, two in each of two classes.

Most simulations were done with a population of 1000 cells across 1100 generations. (In general, results remain qualitatively the same in runs when drug administration is extended over 10^4 generations, and where populations are increased to 10^4 cells (allowing up to 10^5 virions) show the same qualitative

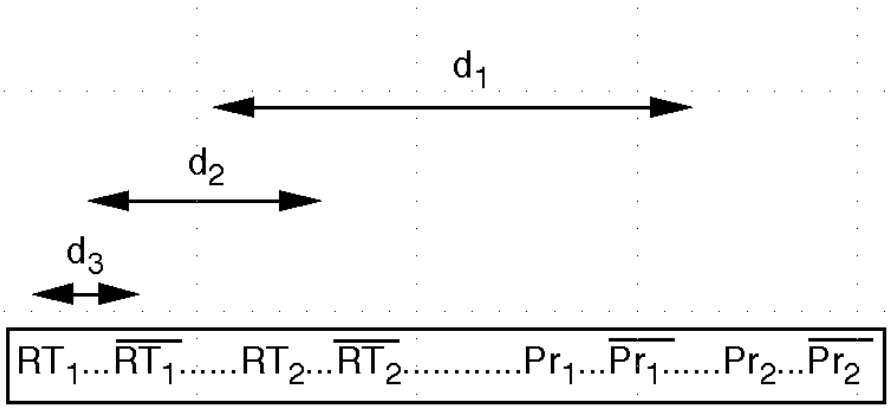


Figure 5: Genomic arrangement

results.) Each cell could be infected by up to 10 virions, and produced 10 virions in each cycle of viral evolution. XOverOccurRate determined the probability that a new virion was formed via recombination across "parent" particles. The relative rates of inter-gene recombination were [0.1,0.2,0.1,0.4,0.1,0.2,0.1]; i.e, it is twice as likely to segregate bits associated with two drugs from the same class as it is to segregate those corresponding to the resistance and and compensatory bits, and twice again as likely to do so across the two classes of drugs. [1] The per-gene mutation rate was 10^{-4}. Tissue was modeled as a one-dimensional array with periodic boundaries (i.e., wrap-around). A new population of cells was initially subject to a constant (0.1) probability of infection by wild-type virus[2] via serum, followed by tissue-specific migration (modeled as a random walk, length= 5) of virus by neighboring infected cells.

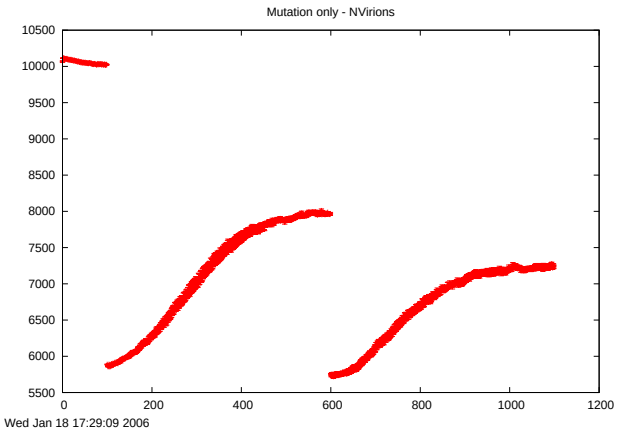
Simulations were begun in a fully infected state, with each cell's complement of virions filled with wild-type virus. A typical drug regime ran this system for 100 generations without drug intervention (to ensure equilibrium), followed by 500 generations of treatment by one drug, followed 500 generations of treatment by a second drug from a different class. Results reported are the average of 10 such runs begun from different random number seeds.[3]

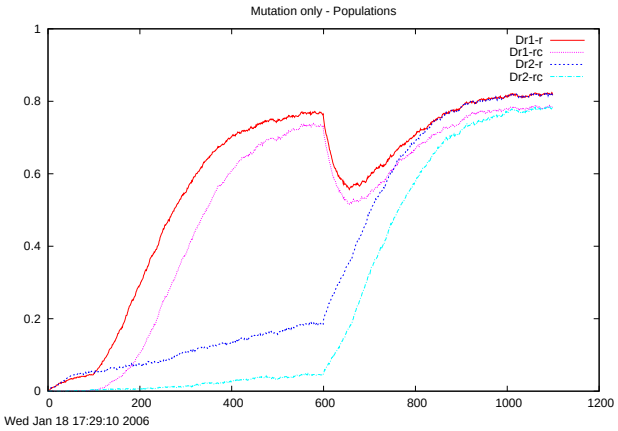
The basic features of the evolutionary progression are shown in the graphs of Figure 6 , showing an initial simulation where XOverOccurRate=0.0, inhibiting recombination and allowing only mutational change. The first graph shows the total number of virions as a function of generational time. Administration of the first drug drops the virion numbers almost in half. Mutants resistant to it

[1] These raw values help to make intuitive sense of relative crossover rates. They must of course be normalized into probabilities.

[2] Note that the assumption that serum infection was always by wild-type vs. mutant virus is an obvious simplification, making results below a *lower bound* on expected mutant levels.

[3] Standard deviation bars are shown for only the NVirion population plots, as they remain very tight in all cases and obscure results otherwise.





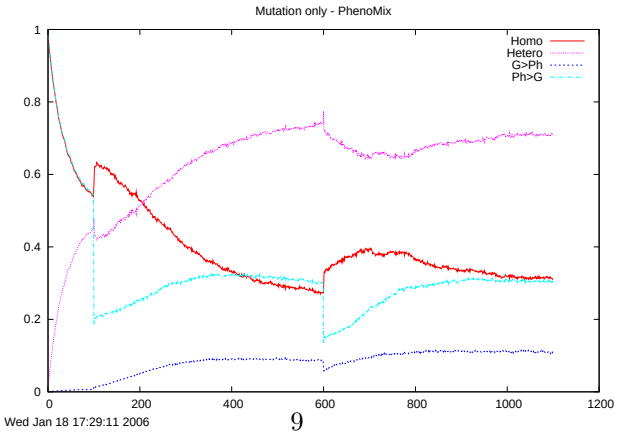


Figure 6: Mutation-only evolution

grow to significant levels after about 100 generations, driving virion populations to approximately three quarters of their original number.

The second graph shows the genetic variability of the virion population, as the proportion of the virion population exhibiting each of the four mutants.[4] For the first 100, untreated generations, two varieties of single (1x) mutants emerge, as consistent with quasi-species drift around wild-type
[Eigen et al., 1989, Domingo and Holland, 1997]. With administration of the first drug, resistance mutations to Drug1 (plotted in red) begin to emerge immediately. At a slower rate, compensatory mutations (pink) come to improve the fitness of approximately two-thirds of the resistant mutants. Note also that levels of mutants resistant to the second (as yet unadministered) Drug2 (dark and light blue) also increase during this period, presumably "hitchhiking" on genomes of Drug1-resistant mutants.

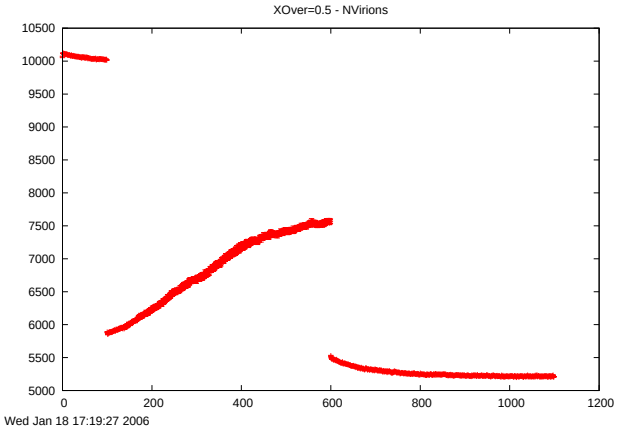
With administration of the second drug, the total number of virions is reduced (further than in response to the first), but again rebounds to almost the same levels. Mutations resistant to Drug2, and then also compensatory to it, quickly join this population until virtually the entire (non-serum infected wild-type) population benefits from both mutation types.

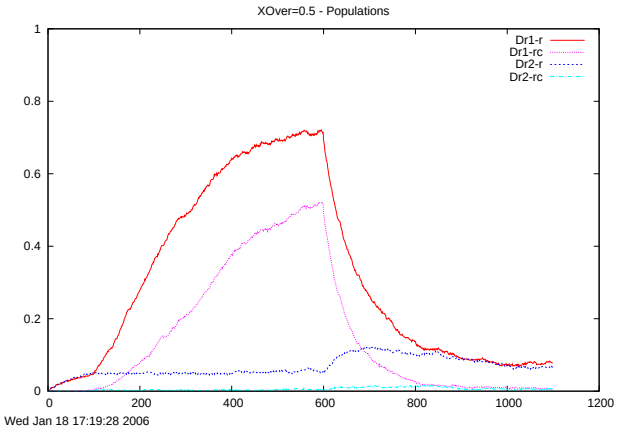
The third graph shows statistics underlying the phenotypic mixing. The first two curves show an initially homogeneous (genotype matches phenotype, plotted in red) population of virions being supplanted by more and more heterogeneous (phenotypically mixed, plotted in pink) ones. A curious spike in the number of homogeneous virions is observed as an initial response to both drugs. . The second pair of curves show the number of selective events where fitness based directly on genotype is greater than that actually associated with phenotype (plotted in dark blue) and *this difference matters* in that the stochastic choice determining whether the virion lives or dies would have been different; the alternative (phenotypic fitness is superior and matters) is also shown (light blue). Heterogeneous virions continue to dominate the population throughout both drug treatments. Phenotypic masking, where phenotypic traits hide the genetic material being inherited, occurs approximately 25% of the time, much more frequently than the converse.

3.1 Allowing recombination

The importance of recombination with respect to phenotypic mixing can be established by setting the recombination rate to 0.5; these results are shown in Figure 7 . The most obvious, immediate consequence is that the number of virions does not exhibit the rapid increase in response to the second drug, but drops to lower levels then after one drug alone (but only to a baseline level, it is not entirely killed).

[4]Note that since the mutations with respect to the two drugs are modeled as independent, their relative levels are unconstrained with respect to one another. Compensatory changes, however, are *contingent upon* the initial resistance mutation, and so can never exceed levels of the associated resistance mutation.





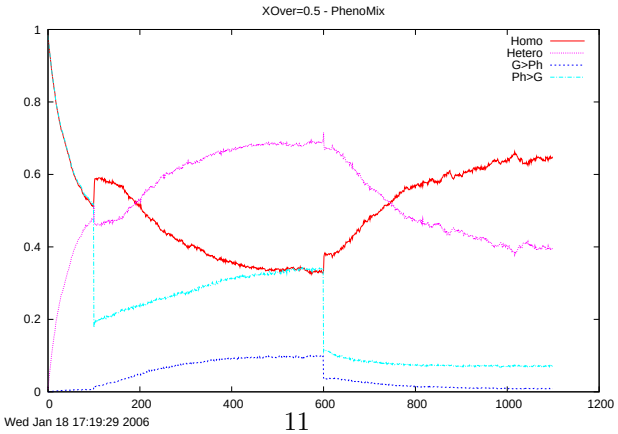


Figure 7: Allowing recombination

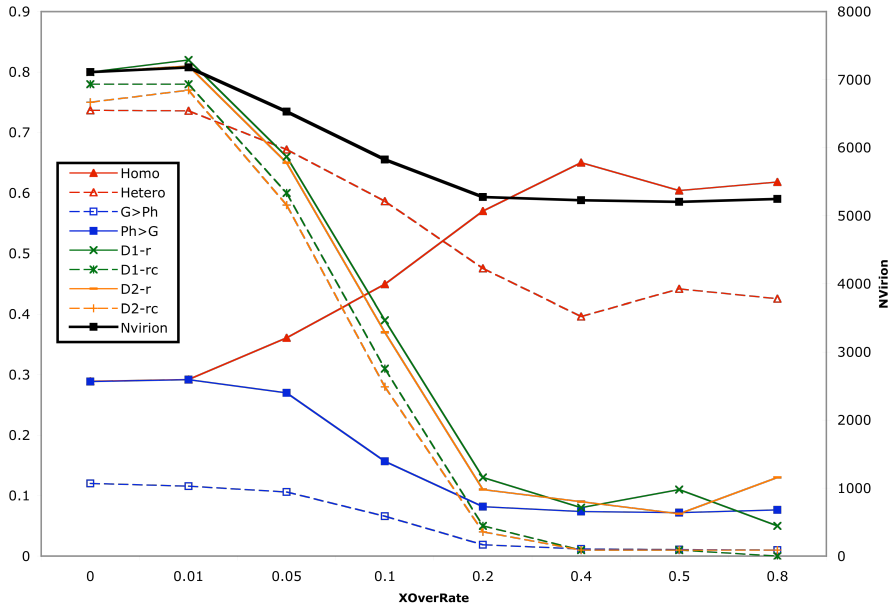


Figure 8: Varying recombination rates

The makeup of this reduced population is also much different: As expected, there is an increase in the number of Drug2 resistant mutants, but it is quite slight; an even smaller number of Drug2 compensatory mutations is observed. Further, Drug1 resistant and compensatory mutants are reduced to levels comparable to the initial untreated state.

More striking still is the different phenotypic mixing response to the second drug: there is now strong selective pressure *against* phenotypic mixing. Homogeneous virions become dominant and phenotypic masking is much reduced.

Figure 8 summarizes experiments showing the final (after all 1100 generations) values for all dependent measures as the recombination rate takes on values other than 0.5. Note that the main effects of recombination remain present even as this value approaches 0.1. Even at these low levels, the dramatic differences observed seem to suggest that *reduced recombination rates would have selective advantage*, despite current estimates which suggest "HIV-1 undergoes recombination at a minimum rate of 2.8 crossovers per genome per cycle." [Zhuang et al., 2002].

4 Summary

The presentation here does not allow space for a full review of the many modeling efforts related to our project. In particular, there has been a long tradition of modeling HIV infection dynamics which depend on differential equation models

(e.g., [Bonhoeffer and Nowak, 1997, Perelson and Nelson, 1999]) quite different from the discrete event modeling used here.

The "mechanistic agent-based stochastic simulation" (MASS) presented by Leonard and Schaffer [Leonard and Schaffer, 2005] uses a discrete event model quite similar to our own. Because their interest is in potential disruptions to the Tat-TAR loop via RNA interference strategies, their model necessarily includes a much more detailed accounting of the viral infection cycle than our "null" phenotype=genotype mapping. Genetic sequences are represented with nucleotide bases rather than the simplified binary genome used in our work. MASS also addresses variation in viral integration sites and gene expression; both of these factors effect relative numbers of progeny produced, a value that is fixed in our model. Finally, phenotypic mixing is not an issue in MASS, since as the TAR loop is a fundamental feature of the viral genome.

As mentioned above, the work most closely related to the experiments reported here is by Bretscher et al. [Bretscher et al., 2004]. As shown in Figure 9, their model uses a selection coefficient depending exclusively on a two-allele model and an epistasis coefficient linking them. Drug fitness is then modeled as a simple function depending only on the selection coefficient with respect to each drug and this epistasis coefficient. Since selective response and potential epistasic effects among various mutants are variables we would like to explore independently, this definition seems unfortunately circular. Their model uses another parameter to control the fraction of single (and therefore certainly homogeneous) vs. double-infected (and therefore potentially heterogeneous) cells; they do not consider super-infection by more than two virions. The model does not allow for structured population effects that can arise via serum diffusion vs. intra-tissue type infection patterns. It also does not attempt to model double-mutant compensatory relations among mutants arising in response to the same drug (in distinction to among mutations providing resistance to two different drugs), nor potential effects of the temporal ordering of various drug regimes.

For all these reasons, and because our group has reliably found advantage to the use of recombination in engineering applications, the simulations reported here were originally hypothesized to find regimes beyond those considered by the Bretscher et al. model in which recombination would confer selective advantage. Instead, our results seem to reinforce the interpretation they offer:

> Our simulations show that for any level of epistasis, increasing superinfection increases the frequency of double mutant provirus before therapy, but decreases its rate of emergence during therapy. This is because phenotypic mixing, that is mixing of the viral proteins produced by distinct proviruses inside one cell, is on average detrimental to the fitter provirus but beneficial to the less-fit provirus. (p. 185)

That we should come to approximately the same conclusions, based on a much different simulation, across a much broader range of conditions, suggests surprising robustness to these conclusions. It also helps to build confidence that our discrete event model conforms in important detail to their analysis, making

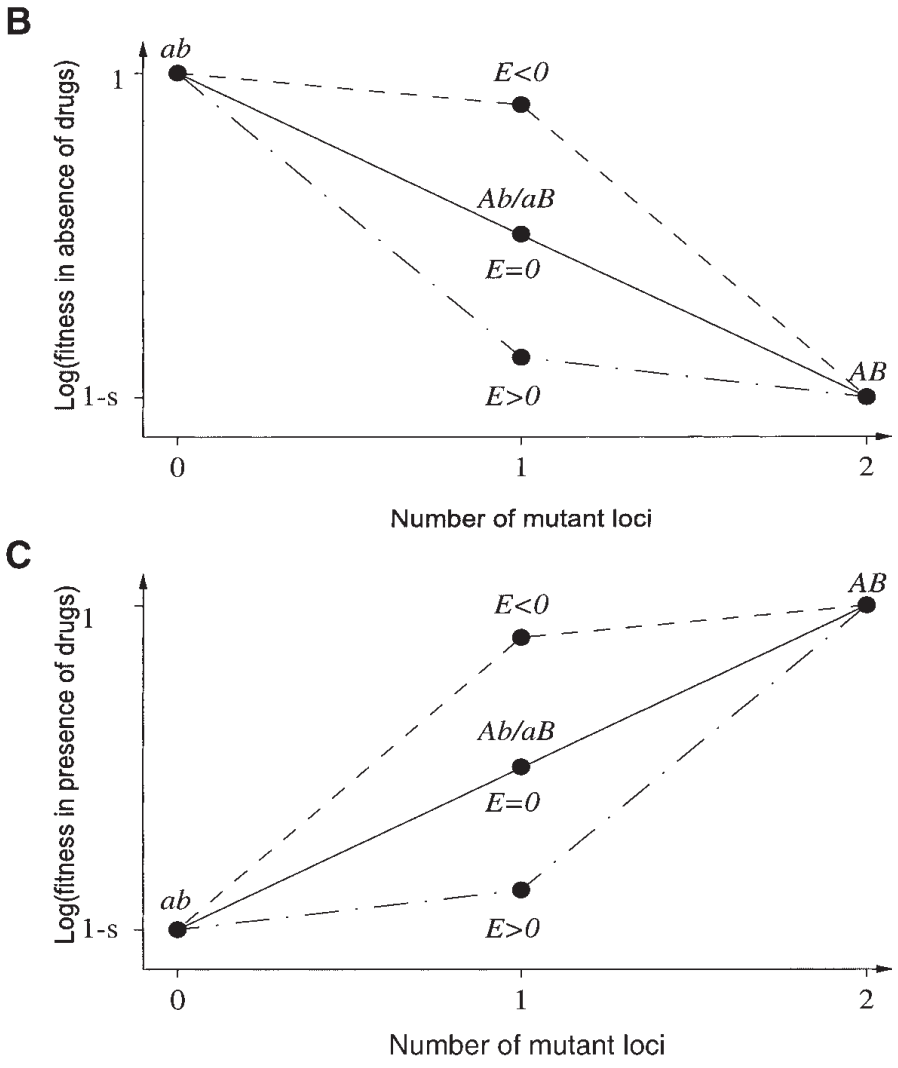


Figure 9: Bretscher et al. model of synergistic/antagonistic epistatic interactions

extensions of our simulation (e.g., to consider other drug regimes, more elaborate tissue typing, etc) into regimes where mathematical analysis is difficult will be worthwhile. It does, however, leave the evolutionary role of recombination in HIV a mystery.

Acknowledgements

The authors gratefully acknowledge long-term collaborations with the Molecular Graphics Lab at The Scripps Research Institute headed by Art Olson and including David Goodsell, Ruth Huey, William Lindstrom and Garrett Morris. We also appreciate the insightful comments of two anonymous reviewers. This work was funded in part under NIH Grant #P01GM48870.

References

[Altenberg, 1995] Altenberg, L. (1995). Genome growth and the evolution of the genotype-phenotype map. In *Evolution and Biocomputation, Computational Models of Evolution*, pages 205–259, London, UK. Springer-Verlag.

[Banzhaf, 1994] Banzhaf, W. (1994). Genotype-phenotype-mapping and neutral variation - a case study in genetic programming. In *PPSN III: Proceedings of the International Conference on Evolutionary Computation. The Third Conference on Parallel Problem Solving from Nature*, pages 322–332, London, UK. Springer-Verlag.

[Bonhoeffer et al., 2004] Bonhoeffer, S., Chappey, C., Parkin, N., Whitcomb, J., and Petropoulos, C. (2004). Evidence for positive epistasis in HIV-1. *Science*, 306(5701):1547–1550.

[Bonhoeffer and Nowak, 1997] Bonhoeffer, S. and Nowak, M. (1997). Pre-existence and emergence of drug resistance in HIV-1 infection. *Proc Biol Sci*, 264(1382):631–637.

[Brenner, 1957] Brenner, S. (1957). Genetic control and phenotypic mixing of the adsorption cofactor requirement in bacteriophages t2 and t. *Virology*, 3(3):560–574.

[Bretscher et al., 2004] Bretscher, M., Althaus, C., Müller, V., and Bonhoeffer, S. (2004). Recombination in HIV and the evolution of drug resistance: for better or for worse? *Bioessays*, 26(2):180–188.

[Burke, 1997] Burke, D. (1997). Recombination in HIV: an important viral evolutionary strategy. *Emerg Infect Dis*, 3(3):253–259.

[Domingo and Holland, 1997] Domingo, E. and Holland, J. (1997). RNA virus mutations and fitness for survival. *Annu Rev Microbiol*, 51:151–178.

[Eigen et al., 1989] Eigen, M., McCaskill, J., and Schuster, P. (1989). The molecular quasispecies. *Adv. Chem. Phys.*, 75:149–263.

[Ewens, 1979] Ewens, W. (1979). *Mathematical population genetics*. Springer Verlag, Berlin.

[Gruau and Whtitley, 1993] Gruau, F. and Whtitley, L. D. (1993). Adding learning to the cellular development of neural networks: Evolution and the baldwin effect. *Evolutionary Computation*, 1(3):213–233.

[Hart et al., 1994] Hart, W. E., Kammeyer, T. E., and Belew, R. K. (1994). The role of development in genetic algorithms. In Whitley, D. and Vose, M., editors, *Foundations of Genetic Algorithms III*. Morgan Kauffman. foga94/abstract.txtfoga94/techreport.

[Holland, 1975] Holland, J. (1975). *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor, MI.

[Kammeyer et al., 1995] Kammeyer, T., Belew, R. K., and Williamson, S. G. (1995). Evolving compare-exchange networks using grammars. *Artificial Life*, 2(2):199–237. sort-net/abstract.txtsort-net/EvoCMPXNets-TR.

[Kellam and Larder, 1995] Kellam, P. and Larder, B. (1995). Retroviral recombination can lead to linkage of reverse transcriptase mutations that confer increased zidovudine resistance. *J Virol*, 69(2):669–674.

[Koza, 1992] Koza, J. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT PRess.

[Leigh Brown and Richman, 1997] Leigh Brown, A. J. and Richman, D. D. (1997). Hiv-1: Gambling on the evolution of drug resistance? *Nature Medicine*, 3(3):268–271.

[Leonard and Schaffer, 2005] Leonard, J. and Schaffer, D. (2005). Antiviral RNAi therapy: emerging approaches for hitting a moving target. *Gene Ther*, 79(3):1645–1654.

[MA et al., 1996] MA, H., PF, S., and W., F. (1996). Smoothness within ruggedness: the role of neutrality in adaptation. *Proc Natl Acad Sci U S A*, 93(1):397–401.

[Moutouh et al., 1996] Moutouh, L., Corbeil, J., and Richman, D. (1996). Recombination leads to the rapid emergence of HIV-1 dually resistant mutants under selective drug pressure. *Proc Natl Acad Sci U S A*, 93(12):6106–6111.

[Perelson and Nelson, 1999] Perelson, A. S. and Nelson, P. W. (1999). Mathematical analysis of hiv-1 dynamics in vivo. *SIAM Rev.*, 41(1):3–44.

[Rouzine et al., 2001] Rouzine, I. M., Rodrigo, A., and Coffin, J. M. (2001). Transition between Stochastic Evolution and Deterministic Evolution in the Presence of Selection: General Theory and Application to Virology. *Microbiol. Mol. Biol. Rev.*, 65(1):151–185.

[Schuster, 1997] Schuster, P. (1997). Landscapes and molecular evolution. *Phys. D*, 107(2-4):351–365.

[Shafer and Schapiro, 2005] Shafer, R. and Schapiro, J. (2005). Drug resistance and antiretroviral drug development. *J Antimicrob Chemother*, 55(6):817–820.

[Zhuang et al., 2002] Zhuang, J., Jetzt, A., Sun, G., Yu, H., Klarmann, G., Ron, Y., Preston, B., and Dougherty, J. (2002). Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. *J. Virology*, 76(22):11273–82.