Ellen Sebastian
Prof. Douglas Brutlag
Biochem 118: Genomics in Medicine

Viral Genomics: Applications to HIV Treatment

The Human Immunodeficiency Virus is as ingenious and fascinating as it is devastating. Although the situation at times seems hopeless- President Clinton's 1997 prediction of a vaccine within the decade seems laughable today- the emerging field of genomics provides glimmers of hope. Since HIV integrates its RNA genome into the human chromosomes, the human and HIV genomes are intimately related. Genomic study of the virus is essential for understanding and combatting it. It is essential that we understand the nuances of HIV's genome, protein products, and interactions with the human genome. This review will first survey current knowledge about the HIV genome. After evaluating current HIV treatment protocols, it will then explore two ways in which genomics improves the efficacy of these protocols: the development of new antiretroviral drugs and genomic resistance testing.

Background: Genomic study of HIV

Current knowledge about the HIV genome is both broad and deep. Unlike an organismal genome, which is always encoded by double-stranded DNA, HIV's genome is stored in single-stranded RNA. The HIV genome is only 9800 base-pairs long and is composed of only 9 genes: *gag*, *pol*, *env*, *tat*, *rev*, *nef*, *bif*, *vpr*, and either *vpu* or *vpx*. Interestingly, these 9 genes encode 15 proteins via cleavage of the *Gag-pol* precursor by viral Protease. The *gag-pol* precursor encodes

Reverse Transcriptase (RT) and Protease, the two most important molecules for HIV's replication and pathogenicity ("P03367 (pol_hv1br," 1988). Study of *pol* and *gag* is therefore especially relevant for the development of new HIV treatments. The HIV genome's shortness and simplicity has allowed it to be widely and thoroughly analyzed using high-throughput RNA analysis techniques, especially with regards to the *gag-pol* system.

In addition to the raw nucleotide sequence, the secondary and tertiary structure of HIV's genome- that is, the three-dimensional steric interactions between nucleotides- and its protein products have also been described in detail. Watts et. al (2009) discovered "previously unrecognized, but readily identifiable and evolutionally conserved, RNA structures" by secondary analysis of the HIV genome. These structures included stem-loops, pseudoknots, and various unstructured motifs. Because these secondary RNA structures are linked to the protein structures they encode, secondary genomic analysis of HIV can promote a deeper understanding of its mechanism and generate new ideas to hinder its operation.

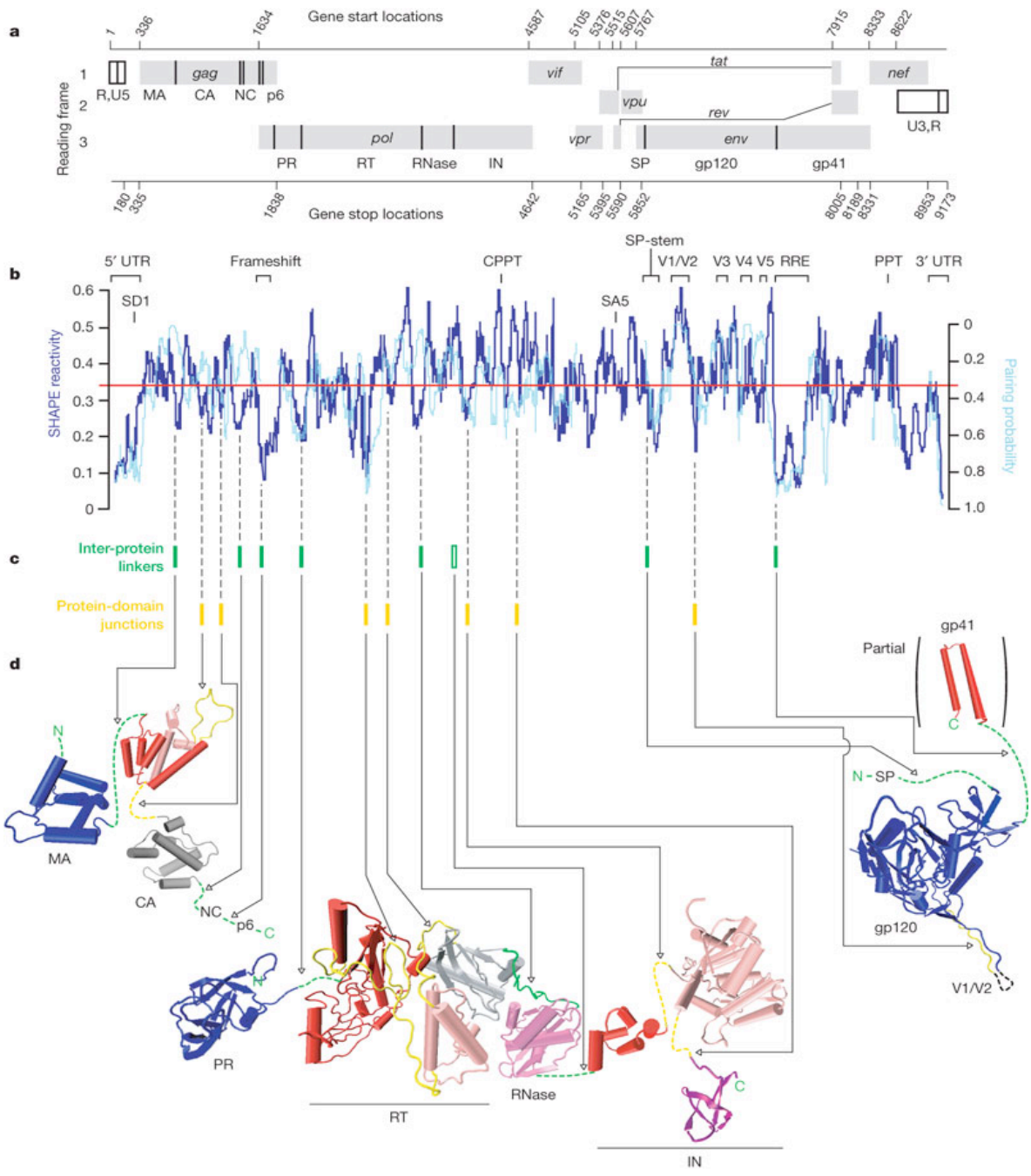

Fig. 1. A map of the ssRNA HIV genome. Unlike a eukaryotic genome, HIV's genome is composed mostly of coding (grey-boxed) regions with only two introns. All nine genes and fifteen proteins are shown; note that the gag-pol precursor includes RT and protease as products. (Watts et. al, 2009)

Clearly, the HIV genome is simpler than any eukaryotic genome. This might lend the impression that genomic study of HIV, and its applications to treatment, is also simple. However, HIV genomics is complicated by two factors: variability and evolution.

HIV's genome is highly variable across different regions of the world as well as between and even within hosts. Genomic sequencing has allowed virologists to classify HIV into two classes, three groups, and dozens of clades (Stebbing & Moyle, 2003). The HIV-1 class is the major contributor to the current HIV epidemic. Unlike most viruses, HIV is capable of genetic recombination; "if the co-packaged SS RBA genomes were derived during infection of a single cell by viruses with different sequences, then recombination during the next cycle of replication produces mosaic viral sequences that may differ from the parental genomes" (Mageridon-Thermet & Shafer, 2010). Genetic recombination contributes to population variability. The variability between HIV strains complicates the development of HIV treatments, analogous to a race "against a whole team of fast runners as opposed to one single fast opponent" (Stebbing & Moyle, 2003).

Another problematic aspect of the HIV genome is its exceptionally fast rate of mutation and evolution. Viruses reproduce very quickly, and viral RNA replication exhibits much lower fidelity than cellular DNA replication because it lacks a proofreading mechanism. In a phenomenon known as intra-host evolution, "Genetic changes in the viral swarm occur rapidly,

particularly when selective pressure is applied when drugs are given" (Stebbing & Moyle, 2003). Over the entire population of HIV virions in a viremic patient's system, point mutations occur 1,000 to 10,000 times per day, or 1 in 100,000 replications (Shafer, 2002). Mutations providing resistance to antiretroviral drugs are strongly selected for and quickly become dominant in a host undergoing treatment. As a result, it would extremely difficult to synthesize a single antiretroviral drug that continues to be effective throughout a patient's life; HIV is winning the molecular arms race. HIV's astronomical mutation rate also prevents the development of an HIV vaccine because an attenuated virus could easily mutate back into a virulent form (Stebbing & Moyle, 2003).

It falls to genomics and bioinformatics to identify the specific mutations involved in drug resistance. These mutations are generally SNPs as opposed to copy number variations, insertions, or deletions (Shafer, 2002). High-throughput sequencing techniques are applied to identify mutations leading to resistance for specific drug classes. For instance, Archer et. al (2000) showed that the amino-acid changes caused by mutations V106a, V179D, and Y181C confer resistance to NNRTIs. Shafer & Schapiro (2008) identify many specific mutations yielding resistance to all six major antiretroviral drug classes. Furthermore, they describe the mechanism in which the mutation alters viral function and allows it to reproduce in the presence of the drug. For example, alterations at the *gag* cleavage site, including the mutations A431V, K436E, and

I437T/V, produce resistance to PIs by allowing protease to cleave the *gag-pol* precursor in a different manner than wild-type protease. As of 2008, over 200 specific mutations related to ARV resistance had been identified.

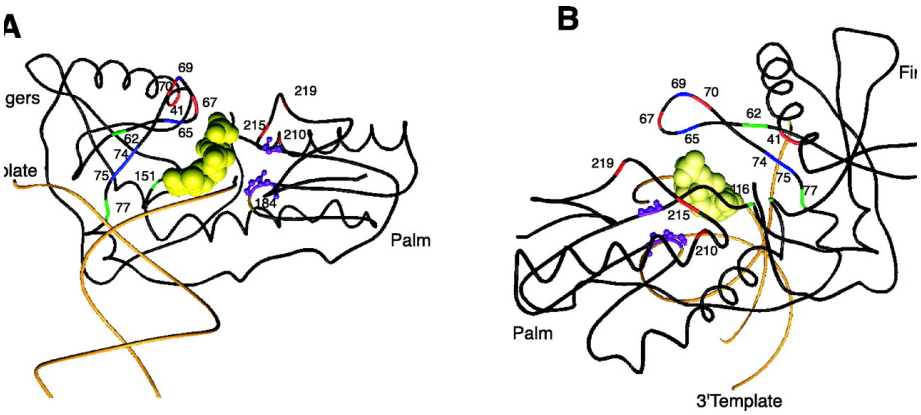

Fig. 2 HIV's RT enzyme labeled with amino-acid changes that promote NRTI resistance in HIV (Shafer, 2002)

	Discriminatory Mutations				Thymidine Analog Mutations (TAMs)						MDR Mutations	
	184	65	74	115	41	67	70	210	215	219	69	151
Cons	M	K	L	Y	M	D	K	L	T	K	T	Q
3TC	**VI**	**R**									**Ins**	**M**
FTC	**VI**	**R**									**Ins**	**M**
ABC	**VI**	**R**	**V**	E	L			W	YF		**Ins**	**M**
ddI	**VI**	**R**	**V**		L			W	YF		**Ins**	**M**
TDF	***	**R**	*	F	L		R	W	YF		**Ins**	**M**
d4T	***	**R**			L	N	R	**W**	**YF**	QE	**Ins**	**M**
ZDV	***	***	*		L	N	R	**W**	**YF**	QE	**Ins**	**M**

Classification: Bold underline: High level phenotypic and/or clinical resistance. Bold: Moderate phenotypic and/or clinical resistance. Plain: Low-level cross-resistance.
Mechanisms: Discriminatory mutations inhibit NRTI incorporation. TAMs promote excision of chain-terminating NRTIs. T69ins is a 2-amino acid insertion which nearly always occurs with multiple TAMS; Q151M usually occurs with V75I, F77L, and F116Y.
M184VI: Although they cause high-level phenotypic resistance to 3TC and FTC, M184VI are not contraindications to 3TC and FTC because M184VI increase TDF, AZT, and d4T susceptibility (***) and decrease viral replication fitness.
Additional NRTI-selected mutations: K65N and K70E/G (similar but weaker effects than K65R), L74I (ddI/ABC without increased AZT/TDF susceptibility), V75TMAS (ddI, d4T), T69D (ddI), E40F, E44DA, D67GE, T69SAING, K70NQT, V118I, H208Y, D218E, L228HR, N348I. Deletions between codons 67 to 70 similar but weaker effects than T69ins (occurs with TAMs or Q151M). T215SCDEIV evolve from T215YF in the absence of NRTIs.

Fig. 4 Summary of mutations yielding resistance to NRTIs (Tang & Shafer, 2012)

HAART and the problem of ARV resistance

The current protocol in HIV treatment is HAART: Highly Active Antiretroviral Therapy. In essence, HAART does not rely on genomics, but genotypic applications dramatically enhance its effectiveness. Treatment with a single antiretroviral drug (ARV) would very quickly lead to resistance and virologic failure. HAART attempts to solve this problem by combining several ARVs in a drug cocktail. Currently there are 24 ARV in six classes (Tang & Shafer 2012). Four classes inhibit the viral enzymes protease, integrase, and reverse transcriptase, and the fusion ihnhibtor and CCR5 inhibitor prevent viral binding to human CD4 cells. In the HIV life cycle, protease is responsible for cleaving the *gag-prot* precursor into its six proteins; integrase inserts the viral DNA prophage into the human genome; and reverse transcriptase transcribes the viral RNA genome into a DNA molecule ("P03367 (pol_hv1br," 1988). Protease inhibitors (PIs) bind to protease and prevent it from cleaving the *gag-prot* precursor, which renders HIV unable to assemble virions. PIs are generally the strongest class of ARV drugs because of their high genetic barrier to resistance, meaning that multiple mutations, as opposed to a single SNP, are required to confer resistance. Reverse Transcriptase Inhibitors (RTIs) can be further divided into Nucleoside RTIs (NRTIs) and Non-nucleoside RTIs (NNRTIs). Through different mechanisms, NRTIs and NNRTIs both inhibit RT to prevent viral replication. NRTIs imitate and compete with nucleosides for incorporation into the DNA chain that RT attempts to extend. When

incorporated in place of a nucleoside, an NRTI prevents further reverse transcription (Margeridon-Thermet & Shafer, 2010). NNRTIs are simply non-competitive enzyme inhibitors. They bind to RT, distorting its shape and preventing it from catalyzing the reverse transcription reaction (Margeridon-Thermet & Shafer, 2010). In order to enter human cells, HIV must bind to one of two human co-receptor proteins: CCR5 or CXCR4. The fusion inhibitor Efurvitide prevents the HIV proteins gp41 and gp120 from binding to human CD4 cells; it is extremely potent but not widely used because of its side effects, including the complete eradication of the host immune system due to the inhibition of CD4 function (Kuritzkes 2009). Generally, Efurvitide is only used in "salvage regimens": cases in which milder cocktails have produced virologic failure (Shafer & Tang, 2012). CCR5 inhibitors also prevent HIV from binding to the surface of CD4 cells, but do so by altering the structure of the CCR5 protein rather than the virus

HAART utilizes a combination of several drugs- always a protease and RTI and often an integrase- to counteract HIV's mutation rate in hope that the combination of several forms of inhibition will raise the genetic barrier. In order to sustain replication, a virus would need to be resistant to *all* components of HAART; the genetic barrier to resistance is prohibitively high. It is extremely unlikely that a virus could simultaneously mutate to acquire resistance to all HAART components. As a result, correctly-applied HAART prevents viremia and AIDS in HIV patients.

HAART is certainly valuable for extending the length and quality of life for HIV patients, but it also exhibits significant disadvantages (Peters & Conway, 2011). First, the practical concerns: HAART requires strict adherence to a complicated medication regimen and is prohibitively expensive for third world countries (which happen to be those most severely handicapped by the AIDS epidemic). Its side effects are severe, ranging from neuropsychiatric to metabolic, and may result in noncompliance (Peters & Conway, 2011). This noncompliance, in turn, enables the virus to mutate and acquire resistance even more easily (Zeller & Kumar, 2011). HAART can actually promote multidrug resistance by providing selective evolutionary pressure: among treated patients, drug resistance is 39 to 53%, in contrast to 5 to 20% when treatment is unavailable (Ceccherini-Silberstein, 2010).

Correctly-applied HAART is relatively effective in preventing viremia, but the problem of antiretroviral resistance remains a concern in the third world. If an ideal combination of drugs is not used, the barrier to resistance is lowered and HAART's potency is severely compromised. In the third world, diagnostic errors and drug unavailability often prevent correct application of HAART (Tang & Shafer, 2012). PIs are the most potent component of HAART, but they are also the most expensive; they are therefore often omitted from HAART in impoverished regions, with devastating consequences. The failure to recognize virologic failure also contributes to the relative prevalence of ARV resistance in the third world (Tang & Shafer, 2012).

Applications of Genomics to HIV treatment

Despite the undeniable improvement that HAART has reaped on the state of HIV treatment, current treatment methodologies are not ideal. We must strive, through further application of genomics, to construct novel approaches to HIV treatment. I will outline two developing applications of genomics: improved antiretrovirals and genotyping resistance testing.

Design of New Antiretrovirals: Genomics in the HIV Pharmaceutical Value Chain

First, studying the HIV genome can allow us to formulate more effective antiretroviral drugs by identifying potential drug targets, which will become increasingly important as HIV refines its resistance to existing ARV (Pomerantz, 2004). Functional genomics can identify "new targets for antiviral agents and downstream targets for therapies to interdict in virus-induced pathogenetic processes" (Pomerantz, 2004). This identification of new viral targets is essential because the failure of existing drugs often results from the absence of an appropriate target. These new ARV may be either small-molecule inhibitors analogous to PIs and RTIs, or they may interfere with host-virus interactions in a way that prevents HIV replication.

Bioinformaticians may aim to analyze HIV's genome in order to synthesize new small-molecule inhibitor ARVs which target its protein products. One example of such analysis concerns the viral gene *vif*. Lecossier et. al (2003) found that *vif* ensures HIV's survival by preventing hypermutation, but demonstrated that human CEM15 can induce *vif* failure. This

inhibitory interaction between CEM15 and *vif* may provide a model to design antiretrovirals that attack *vif*, which is not currently the target of any of the six classes of ARV.

Secondly, new antivirals may target host-virus interactions. Current research attempts to expand the class of CCR5 inhibitors, which disrupt HIV's entry into CD4 cells by binding to or 'blockading' human CCR5 receptors. For example, Ibalizumab is a monoclonal antibody that prevents HIV's gp120 binding protein from accessing CCR5; unlike Efurvitide, it does not impair normal CD4 cell function so does not incapacitate the host immune system (Kuritzkes 2009). Because CCR5 antagonists target a host cell protein instead of a viral protein, they are slightly less sensitive to ARV resistance than small-molecule inhibitors. However, the possibility of resistance to CCR5 antagonists remains because HIV can evolve a different binding mechanism that allows entry to the CD4 cell despite the alteration of CCR5 (Kuritzkes 2009). Worryingly, Westby (2006) found that following treatment with the CCR5 antagonist Maraviroc, HIV evolved to bind to the CXCR4 receptor instead, for which no inhibitor exists. Nevertheless, CCR5 inhibitors are the future of antiretroviral drugs.

Designing ideal HAART cocktails: Genotypic drug resistance testing

Genomics has worked in concert with HAART to counteract its disadvantages and enhance quality and length of life for HIV patients. One example of the application of viral genomics to existing HIV treatment is genotypic drug resistance testing, which augments

HAART's personalization and selectivity. If applied blindly without knowledge about the HIV infection in question, HAART would be ineffective because the virus could be already resistant to the drugs in the cocktail. Instead, HAART regimens should be personalized on a case-by-case basis according to virus's resistance factors. Because drug resistance is derived from mutations in the HIV genes encoding drug targets, genomic analysis is useful for identifying not only the presence, but also the magnitude of a particular virus strain's resistance to the drugs under consideration (Shafer, 2002). Before a patient begins an antiretroviral regimen, a sample is taken of the specific HIV virus found in the patient's blood plasma. After PCR magnification (Shafer, 2002), the virus's genome is sequenced using Sanger sequencing (Ceccherini-Silberstein, 2010). This method is used in concert with viral culture in the presence of antiretrovirals to determine resistance (Shafer, 2002). However, the genotypic method is preferable to the virologic method because it is faster, cheaper, and more accurate (Ceccherini-Silberstein, 2010). After analyzing the results of genotypic drug resistance testing, physicians can then select HAART components that will be effective against the HIV strain in question. Genotypic drug resistance testing has been shown to increase the efficacy of HAART in clinical settings (Shafer, 2002). Genomics, therefore, complements the personalization of current HIV treatments. Ceccherini (2010) and Shafer (2002) both recommend the expansion of genotypic drug resistance testing to all HIV cases.

Genotypic drug resistance testing is incredibly useful but also demanding; it is a challenge for physicians to interpret genotypic results and design a HAART cocktail. The field of bioinformatics has helped to ameliorate this issue by creating large databases of viral genomic information. In order to facilitate the personalization of HAART regimens, Robert W. Shafer at Stanford University developed HIVdb: "a freely available online genotypic resistance interpretation system… to help clinicians and laboratories interpret HIV-1 genotypic resistance tests" (Tang, Lui & Shafer, 2012). The database quantifies antiretroviral-resistant mutations present in an HIV sample and assigns scores for each mutation, resulting in an estimate of the extent to which that HIV strain is resistant to each type of antitretroviral drug. Its algorithm also compensates for interactions between mutations, resulting in a broader, more accurate picture of the virus's resistance capacities. Crucially, it also provides recommendations for specific drugs that should be included in the HAART cocktail. Users may enter either the amino-acid sequence or a list of mutations corresponding to their sample's genome. The database returns data about the sample's relative levels of resistance to different ARV and recommendations for HAART cocktails. The extensiveness, accuracy and convenience of such a database testifies to the power of genomics in a practical clinical setting.

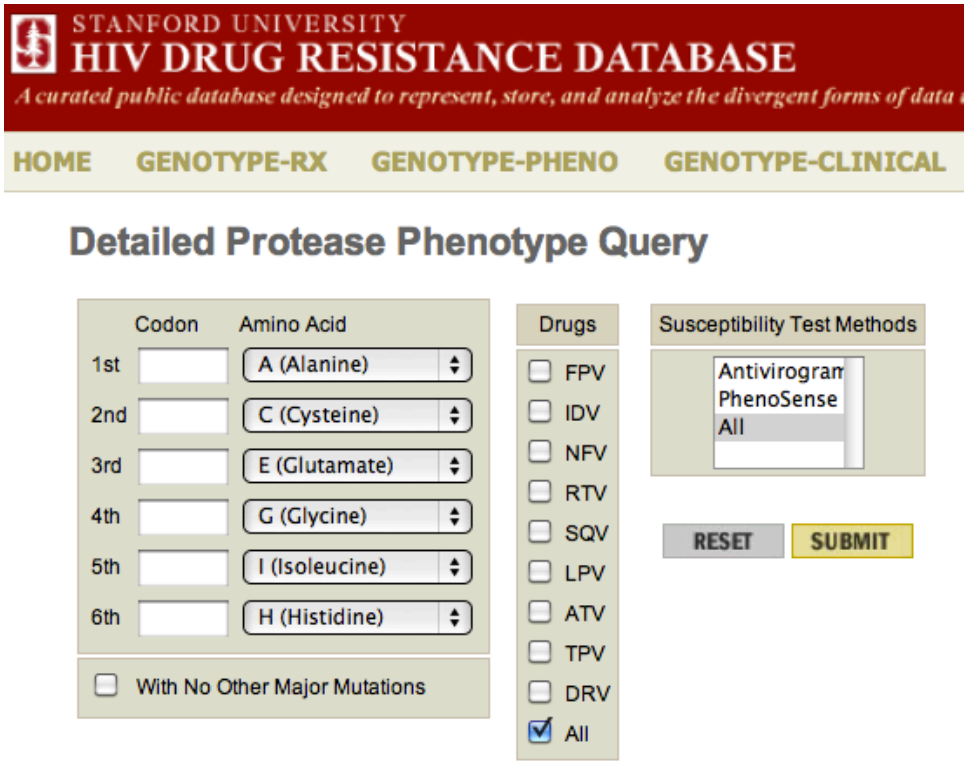

fig. 4. The HIVdb user interface for analyzing an HIV sample's resistance to the 8 PIs listed under "drugs." The user enters up to 6 amino acids found at certain codons in the sample. The database returns a graphic illustrating these amino acids' effects on the sample's resistance to the drugs selected. This is one of many database search tools available at hivdb.stanford.edu.

As part of a larger trend toward the application of genomics toward medicine's most pressing issues, genomics is an essential component in the battle against the HIV epidemic. Already we have seen promising results: new ARVs such CCR5 inhibitors appear to be effective, and genotypic resistance testing compensates for some of HAART's shortcomings. Although HIV's flexibility and lethality make it a formidable opponent, genomic approaches give us an edge toward eventually defeating it.

REFERENCES

Archer, R. H., Dykes, C., Gerondelis, P., Lloyd, A., Fay, P., Reichman, R. C., Bambara , R. A., & Demeter, L. M. (2000). Mutants of human immunodeficiency virus type 1 (hiv-1) reverse transcriptase resistant to nonnucleoside reverse transcriptase inhibitors demonstrate altered rates of rnase h cleavage that correlate with hiv-1 replication fitness in cell culture. Journal of Virology, 75(18), 8390-8401. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC116350/

Ceccherini-Silberstein, F., Cento, V., Calvez, V., & Perno, C. F. (2010). The use of human immunodeficiency virus resistance tests in clinical practice. Clinical Microbiology and Infection,16(10), 1511-1517. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1469-0691.2010.03353.x/full

Lecossier, D., Bouchonnet, F., Clavel, F., & Hance, A. J. (2003). Hypermutation of HIV-1 DNA in the absence of the Vif protein. http://onlinelibrary.wiley.com/doi/10.1111/j.1469-0691.2010.03353.x/full. Science, 300(5622), 1112. doi: 10.1126

P03367 (pol_hv1br) reviewed, uniprotkb/swiss-prot. (1988, November 1). Retrieved from http://www.uniprot.org/uniprot/P03367

Margeridon-Thermet, S., & Shafer, R. W. (2010). Comparison of the mechanisms of drug resistance among hiv, hepatitis b, and hepatitis c. *Viruses*, *2*, 2696-2739. Retrieved from https://docs.google.com/viewer?a=v&pid=gmail&attid=0.1&thid=135e1ca5a547fca0&mt=application/pdf&url=https://mail.google.com/mail/?ui%3D2%26ik%3D115aba47fa%26view%3Datt%26th%3D135e1ca5a547fca0%26attid%3D0.1%26disp%3Dsafe%26zw&sig=AHIEtbRORORm0DAxhT3TVD3sEv_dZwJJ6Pg

Peters, B. S., & Conway, K. (2011). Therapy for HIV past, present, and future. Advances in Dental Research,23(1), 23-27. Retrieved from http://adr.sagepub.com/content/23/1/23.full

Pomerantz, R. J. (2004). Hiv-1 infection and genomics: Sorting out the complexity. Journal of Infectious Diseases, 189(4), 567-571. Retrieved from http://jid.oxfordjournals.org/content/189/4/567.full

Shafer, R. W., & Schapiro, J. M. (2008). Hiv-1 drug resistance mutations: an updated framework for the second decade of haart. *AIDS Reviews*, *10*(2), 67-84. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2547476/?tool=pubmed

Shafer, R. W. (2002). Genotypic testing for human immunodeficiency virus type 1 drug resistance. *Clinical Microbiology Reviews*, *15*(2), 247-277. Retrieved from http://cmr.asm.org/content/15/2/247.full

Stebbing, J., & Moyle, G. (2003). The clades of hiv: Their origins and clinical significance. AIDS Reviews, 5, 205-213. Retrieved from http://www.aidsreviews.com/files/2003_05_4_205-213.pdf

Tang, M. W., Liu, T. F., & Shafer, R. W. (2012). The hivdb system for hiv-1 genotypic resistance interpretation. Intervirology, 55, 98-101. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22286876

Tang, M.W., & Shafer, R.W. (2012) Impact of HIV-1 Antiretroviral Resistance on Treatment Options: Selecting the Right Regimen for Each Patient. Unpublished article.

Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess, J. W., Swanstrom, R., Burch, C. L., & Weeks, K. M. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. Nature,460, 711-716. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2724670/

Westby, M., Lewis, M., Whitcomb, J., Youle, M., Pozniak, A. L., James, I. T., Jenkins, J.M., Perros, M., & van der Ryst, E. (2006). Emergence of cxcr4-using human immunodeficiency virus type 1 (hiv-1) variants in a minority of hiv-1-infected patients following treatment with the ccr5 antagonist maraviroc is from a pretreatment cxcr4-

using virus reservoir. *Journal of Virology*, *80*(10), 4909-20. Retrieved from

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1472081/?tool=pubmed