

**Hepatitis C Virus Sequencing:
Viral evolution, immune recognition and vaccine development.**

A white paper for consideration by the NIAID Microbial Sequencing Program

Todd M. Allen, Ph.D.
Assistant Professor
Partners AIDS Research Center
Massachusetts General Hospital
Harvard Medical School
tallen2@partners.org

and

Bruce D. Walker, M.D.,
Professor of Medicine
Harvard Medical School
Massachusetts General Hospital-East
bwalker@partners.org

I Executive Summary

The hepatitis C virus (HCV) is a major cause of liver cirrhosis, liver failure and hepatocellular carcinoma, infecting an estimated 170 million individuals worldwide ¹. Furthermore, HCV has emerged as an important cause of morbidity and mortality in HIV-infected subjects ^{2,3}. While drug treatments are available, clear limitations in access and efficacy require concerted efforts to develop effective prophylactic and therapeutic vaccines.

The inherent sequence diversity of HCV, however, represents an enormous challenge facing development of an effective HCV vaccine. Recent studies in HIV suggest that evolution of highly variable pathogens is not random, but highly influenced by evasion of host immune pressures ^{4,5}. Understanding the forces shaping HIV sequence diversity is helping to predict the evolution of HIV and to identify immune responses capable of exerting these selective pressures, issues central to the design of an effective vaccine. Recent data in HCV suggests very similar forces are at work to shape HCV sequence diversity ⁶, but little work has been done to date.

This pilot study represents the first step of identifying the extent to which CD8 T cell responses are driving sequence variation of HCV. We propose to sequence entire HCV genomes (~9kb) from 500 North American (genotype 1a) strains. To maximize the power of these studies and leverage existing resources, samples will be derived from patients in whom full HLA class I typing (A, B, C loci) will be available. These sequence data will identify numerous HLA-associated polymorphisms in HCV at the population level and the role of immune pressures in shaping both regional and global HCV diversity. An understanding of the role of immune pressures in shaping both regional and global HCV diversity is critical in predicting current and future evolution of the virus and thus in guiding vaccine antigen selection.

II Justification

The hepatitis C virus (HCV) is a major cause of liver cirrhosis, liver failure and hepatocellular carcinoma, infecting an estimated 170 million individuals worldwide ¹. While drug treatments are available, only about half of those treated are able to clear the infection. More importantly, in most parts of the world where infection levels are the highest treatment is either unavailable or prohibitively expensive. Recent studies suggest that immune control of HCV is possible ^{7,8}. Specifically, CD8⁺ killer T cells play a key role in governing the outcome of acute HCV infection in both humans and chimpanzees ^{9,10}. Therefore, renewed efforts are needed to develop effective prophylactic and therapeutic vaccines capable of engendering strong CD8⁺ T cell responses.

However, the inherent sequence diversity of HCV, not only within an individual host but in particular on a global scale, represents an enormous challenge facing development of an effective HCV vaccine. Recent studies in HIV suggest that evolution of highly variable pathogens is not random, but highly influenced by evasion of host immune pressures. Population sequencing of HIV together with high resolution HLA typing has revealed accumulation of mutations in HIV associated with expression of specific class I alleles ⁴, and together with other studies provides strong evidence that immune selection can occur in the face of adaptive CD8⁺ T cell responses ¹¹.

¹³. Moreover, in a longitudinal study of acute HIV infection over 65% of mutations arising in HIV-infected subjects could be attributed to CD8 T cell selective pressures (Allen et al, MS in preparation). These data clearly indicate that a majority of the sequence diversity of HIV-1, even at the population level, can be attributed to evasion of immune selection pressures.

With respect to vaccine design, these analyses are beginning to reveal that in the setting of highly polymorphic pathogens immune pressures eventually drive some escape mutations to fixation. This results in some regions of the virus no longer being immunogenic to individuals expressing particular HLA molecules ⁴. These data are even beginning to reveal population specific polymorphisms in HIV suggesting the loss of some immunogenic regions of the virus in one population expressing high levels of a particular HLA allele, while this region remains intact in another population expressing only low levels of the allele. In conjunction with a study suggesting that individuals in a population expressing rare HLA alleles (against which the virus has not yet adapted) may better control HIV infections, has a profound impact upon the selection of specific antigens for inclusion in vaccines ¹⁴.

Whereas viral escape from CD8 responses has been illustrated in the chimpanzee model of HCV infection, the effect of CD8 selection pressure on viral evolution and containment in acute HCV infection in humans remains unclear. Recent data in human HCV infection now supports a role for cellular CD8 T cell responses in shaping HCV sequence diversity within an individual ⁶. However, this study focused on a single immunodominant CD8 T cell response in the NS3 protein. Little is known regarding the broader extent of selective pressures impacting HCV evolution, especially at the population level.

A similarly important determinant can be derived from viral sequence data, that is the relative pressures induced by CD8 T cell responses. Therefore, faster or more frequently arising HLA-associated mutations would presumably be associated with strong CD8 T cell responses. Indeed, high avidity CD8 T cell responses have been associated with rapid escape in the SIV-infected rhesus macaque model of AIDS ¹⁵. Therefore, viral diversity can be used as a surrogate marker of the strength of a CD8 T cell response. Conversely, the rate at which HLA-associated mutations arise may be related to functional constraints within the virus, with more conserved regions of a virus less likely to support evolution. Similar data is available in HIV to support this hypothesis (Allen et al, MS in preparation), as does the observation that CD8 escape mutations will revert upon transmission to a new host lacking the selecting immune pressure ^{5,16}. Taken together, these data support that both the immune responses and the virus are impacting when and where sequence evolution arises. By extension, these data are providing unique insights into the impact of immune responses upon highly variable pathogens such as HCV, HIV, and SIV, and aid in highlighting highly effective CD8 T cell responses as well as regions of these pathogens substantially refractive to sequence evolution ^{5,16}. Each of these issues is central to the selection of which CD8 T cell responses should be engendered by a vaccine hoping to contain or eliminate infections by highly variable pathogens.

While the above studies clearly indicate the impact that host immune responses have on evolution of highly variable pathogens, a recent study now provides data to suggest an impact of the pathogen upon the host's highly polymorphic HLA class I loci. A study comparing HIV-specific immune responses and sequence polymorphisms in North American clade B viruses

verses South African clade C viruses concluded that HLA-B gene frequencies in the population are those most likely to be influenced by HIV disease, consistent with the observation that B alleles are evolving more rapidly than A alleles¹⁷. These data suggest that over the longer course of an epidemic populations may adapt, with HLA alleles best associated with control of the pathogen most likely to expand within a population.

In this study we propose to sequence HCV isolates from 500 HCV-infected subjects derived from North America, predominantly genotype 1a. All individuals will have HLA class I typing data available. Preliminary data from 46 completed genomes has already identified nearly a dozen HLA-associated sequence polymorphisms in HCV, specific to alleles including HLA-A1, -A23, -B8, and -B55. These preliminary data suggest that even modest increases in the size of this data set will reveal dozens of additional associations, including those specific to more rare HLA alleles. These data will provide a better understanding of the extent to which HLA-associated selective pressure are shaping the evolution across the HCV genome, pinpointing regions targeted by highly effective CD8 T cell responses capable of inducing variation.

The scope of this study can then rapidly be expanded to genotypes 1b as well as genotypes 2-6. This will allow us to begin to address genotype- and population-specific polymorphisms in HCV and begin to address the role of particular HLA alleles in the control of HCV¹⁸.

III Current genomic resources

Within the Los Alamos HCV Sequence Database (<http://hcv.lanl.gov/content/hcv-db/index>) there exist a total of 144 full HCV genomes which have been sequenced. Only 12 of these are genotype 1a, 87 genotype 1b, and the remainder distributed across the other genotypes 2-6. Therefore, many strains of HCV have already been sequenced from populations from all around the world. Unfortunately, in the vast majority of these cases the HLA class I alleles of the infected subject were never identified. Similarly, peripheral blood mononuclear cells (PBMC) and additional plasma samples are rarely banked. Therefore, while these sequences are valuable in terms of our understanding of circulating strains of HCV, their potential is limited by these deficiencies.

We propose to sequence strains of HCV from which the source patients' HLA will be available, and from which additional PBMC and plasma samples will be available upon which to conduct additional immunological and virological assays central towards addressing the issue of immune control of HCV. Furthermore, with the expected capacity to sequence HCV we will be able to increase the dataset for many of the genotypes for which fewer than 10 full genomes exist (i.e. genotypes 3, 4, 5, and 6).

We are aware of, and strongly support, the proposal from LANL for additional HCV sequencing. That project will generate an important resource for worldwide research on HCV by improving the classification of HCV in all genotypes. Our approach, which in this pilot is focused on the current health threat in the United States, could subsequently be expanded to address other genotypes.

IV Community depth of interest/demand for genome sequence

Similar efforts in HIV have yielded very important insights into the evolution of this virus⁴ providing a valuable resource to the HIV research community. Therefore a similar resource in HCV is warranted and there is significant local and international interest in this sequencing project. We have consulted with numerous groups in the HCV field regarding interest in these data, including the best approach with which to conduct these studies and the viral strains which would best serve these efforts. Continued discussions with these individuals will help continue to guide these efforts. These scientists include:

Stuart Ray	(Johns Hopkins University)
Chris Walker	(Ohio State University)
Paul Klenerman	(Peter Medawar Center, Oxford University)
David Thomas	(Johns Hopkins University)
Silvana Gaudieri	(Murdoch University, Perth Australia)
Simon Mallal	(Murdoch University, Perth Australia)
Georg Lauer	(Massachusetts General Hospital)

The sequence data generated from this project will be made widely available to the HCV community through data deposition of the sequence traces at the NCBI Trace Archive, as well as the complete viral genome sequences at GenBank and the Los Alamos National Laboratory HCV Sequence Database (www.hiv.lanl.gov). Given that only 12 full length genotype 1a sequences currently exist in the LANL database, our initial focus on genotype 1a sequences will have an added benefit to the community in helping to document the sequence diversity of the genotype.

V Sequencing Project Mgt.

Names and roles of key collaborators.

Todd M. Allen, Ph.D.	10% effort
Assistant Professor	
Partners AIDS Research Center	
Massachusetts General Hospital	
Harvard Medical School	

Bruce D. Walker, MD	Advisory role
Professor of Medicine	
Partners AIDS Research Center	
Massachusetts General Hospital	
Harvard Medical School	

VI Rationale for utility of these genome sequences

A. Recognizing selection through sequence comparisons.

It is already clear that the rate of change of nucleotides is not constant over the HCV genome. However, no resource yet clearly represents either the range of potential diversity across the entire HCV genome or the rate of change at any region. This comparative sequencing project will identify the location and frequency of nucleotide changes across the HCV genome. It will thereby pinpoint regions exhibiting increased rate of change, i.e., more frequent changes arising in independent lineages than seen in neighboring regions. Correlating these signatures of selective pressure with the HLA typing information will provide a wealth of critical information about the interaction of the virus with the host immune system.

B. Strains/sources of DNA/patient populations

- i) Viral strains will be derived from study subjects enrolled at the Massachusetts General and Shattuck Hospitals in Boston, as well as through ongoing collaborations at the Oswaldo Cruz Institute in Rio de Janeiro, Brazil.
- ii) These strains will primarily compose of genotypes 1a (and 1b).

VII Rationale for sequencing strategy

A. Genome size

The size of the HCV genome is roughly 9kb. Full length sequencing of HCV is deemed necessary as very little work has been accomplished to identify specific regions of HCV preferentially targeted by host immune responses.

Bruce Birren 10/29/04 2:40 PM
Formatted: Bullets and Numbering

B. Finished vs. draft sequencing

As identification of nucleotide changes are critical to the analysis, high quality sequence is required to identify real differences rather than sequencing errors. This will be achieved through the sequencing of short overlapping PCR fragments as described below.

Bruce Birren 10/29/04 2:40 PM
Formatted: Bullets and Numbering

C. Sequencing Strategy

- i) Our sequencing approach is currently based on the utilization of a set of 40 PCR primer pairs overlapping the entire HCV genome. A nested PCR approach was deemed most appropriate to yield sufficient amplification based on variable viral loads.
- ii) Our approach relies on population sequencing of bulk viral RNA. In approximately 1% of residues mixed bases are expected, reflecting mixed populations of quasispecies. This approach, therefore, does not yield sequence data from an individual clone, as this could otherwise misrepresent the frequency

of individual sequence polymorphisms in the quasispecies. Therefore, sequencing of individual full length clones would not provide the necessary data to accurately identify HLA-associated polymorphisms.

- iii) PCR on viral samples will be performed using existing protocols in the Investigator's lab at MGH. PCR products will be provided to the Broad Institute MSC for high throughput sequencing. Optimization of all process steps will be coordinated jointly between staff at MGH and the Broad Institute.

VIII Other funding sources.

None for the scale of the project described.

IX References

1. Lauer, G.M. & Walker, B.D. Hepatitis C virus infection. *N Engl J Med* **345**, 41-52 (2001).
2. Graham, C.S. et al. Influence of human immunodeficiency virus infection on the course of hepatitis C virus infection: a meta-analysis. *Clin Infect Dis* **33**, 562-9 (2001).
3. Martinez-Sierra, C. et al. Progression of chronic hepatitis C to liver fibrosis and cirrhosis in patients coinfecting with hepatitis C virus and human immunodeficiency virus. *Clin Infect Dis* **36**, 491-8 (2003).
4. Moore, C.B. et al. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* **296**, 1439-43. (2002).
5. Leslie, A.J. et al. HIV evolution: CTL escape mutation and reversion after transmission. *Nat Med* **10**, 282-9 (2004).
6. Timm, J. et al. CTL Escape and Reversion in Acute HCV Infection. *Journal of Experimental Medicine* **In Press**(2004).
7. Lechner, F. et al. Analysis of successful immune responses in persons infected with hepatitis C virus. *J Exp Med* **191**, 1499-512. (2000).
8. Gruner, N.H. et al. Association of hepatitis C virus-specific CD8+ T cells with viral clearance in acute hepatitis C. *J Infect Dis* **181**, 1528-36 (2000).
9. Shoukry, N.H., Cawthon, A.G. & Walker, C.M. Cell-Mediated Immunity and the Outcome of Hepatitis C Virus Infection. *Annu Rev Microbiol* (2004).
10. Shoukry, N.H. et al. Memory CD8+ T cells are required for protection from persistent hepatitis C virus infection. *J Exp Med* **197**, 1645-55 (2003).
11. Phillips, R.E. et al. Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature* **354**, 453-9 (1991).
12. Price, D.A. et al. Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. *Proc Natl Acad Sci U S A* **94**, 1890-5 (1997).
13. Goulder, P.J.R. et al. Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nature Medicine* **3**, 212-217 (1997).
14. Trachtenberg, E. et al. Advantage of rare HLA supertype in HIV disease progression. *Nat Med* **9**, 928-35 (2003).
15. O'Connor, D.H. et al. Acute phase CTL escape is a hallmark of SIV infection. *Nature Medicine* **8**, 493-499 (2002).

16. Allen, T.M. et al. Selection, transmission, and reversion of an antigen-processing cytotoxic T-lymphocyte escape mutation in human immunodeficiency virus type 1 infection. *J Virol* **78**, 7069-78 (2004).
17. Kiepiela, P. et al. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature In Press*(2004).
18. McKiernan, S.M. et al. Distinct MHC class I and II alleles are associated with hepatitis C viral clearance, originating from a single source. *Hepatology* **40**, 108-14 (2004).
19. Simmonds, P. Genetic diversity and evolution of hepatitis C virus - 15 years on. *J Gen Virol* **85**, 3173-88 (2004).