Title: Molecular Medicine for the 21st Century: A Computational Basis for Design and Critique of Vaccines and Therapeutics

Author(s): A. Lapedes, T-13
G. Myers, T-10
A. Perelson, T-10
R. Farber, T-13
C. Tung, T-10

Submitted to: DOE Office of Scientific and Technical Information (OSTI)

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.
DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.
Molecular Medicine for the 21st Century: A Computational Basis for Design and Critique of Vaccines and Therapeutics

Alan Lapedes*, Gerry Myers, Alan Perelson, Robert Farber, and ChangShung Tung
Theoretical Division, Los Alamos National Laboratory

Abstract
This is the final report of a three-year, Laboratory Directed Research and Development (LDRD) project at the Los Alamos National Laboratory (LANL). The kissing-loop RNA motif is thought to modulate different steps in the retroviral life cycle. Using an RNA folding protocol developed under this project, we are studying the folding of the potential kissing-loop motif in HIV. Using only limited information on base-pairings, and the constraints based on the connectivities, we are modeling the structure and the flexibility of this particular structural motif and have developed three predicted structures with low conformational energy. Following on our earlier work which analyzed correlated mutations in the V3 loop of HIV, we have developed an evolutionary model which incorporates non-independent mutations at different sequence positions in model sequences. This model takes into account the effects of the phylogenetic tree on the analysis of mutations, and shows a higher correspondence of correlated positions with structurally adjacent positions in the model than previous analyses. Additional research under this project has dealt with analyzing the dynamics of HIV in vivo. Using some simple models and patient data, we were able to establish for the first time some quantitative estimates of HIV dynamics in vivo. We found that 10 billion viral particles are produced and cleared each day in a typical mid-stage AIDS patient. Further, we found that the vast majority of the virus (99%) is produced by cells that only live about 1.5 days. These estimates suggest that if HIV replication can be completely stopped by potent drug combinations, then HIV and all cells infected by HIV may be able to be eliminated from the body after two to three years of completely effective treatment.

1. Background and Research Objectives

The rationale for this project is the crisis in health care that is a consequence of newly evolved pathogens, and the enhanced transmission of infectious diseases. The extraordinarily rapid spread of HIV (human immunodeficiency virus), resulting in enormous human suffering and devastation, is a disease in which the Laboratory has already made a major contribution via the HIV Sequence Database. The emergence of

*Principal Investigator, E-mail: asl@t13.lanl.gov
persistent drug resistant strains of tuberculosis, malaria, and other infectious pathogens pose problems that may well prove to be as devastating as AIDS. The gravity of the situation has recently been described in numerous articles in the popular press as well as technical journals. To quote a Science lead editorial "...this is true biological warfare, in which new drugs designed by humans will become obsolete through mutations, only to be replaced by new human drugs and new mutations in a seesaw battle... a long struggle with a premium on basic research to improve our stratagems and applied research to develop new magic bullets is clearly the prognosis for the future."

The objective is to develop a comprehensive, integrated program to bring the great wealth of experimental data currently being generated into a comprehensible framework which would allow reasoned selection of vaccine candidates and therapeutic agents, rapid evaluation of changes in pathogens displaying drug resistance, and new means to overcome this resistance.

2. Importance to LANL's Science and Technology Base and National R&D Needs

The ability to respond effectively to such major health threats will depend largely upon the ability to design effective and inexpensive vaccines and therapeutic agents. Vast amounts of complex biomedical information relevant to the problem of vaccine and therapeutic design are being generated by modern experimental techniques and deposited in national and international databases. The HIV Sequence Database at Los Alamos, which is the international repository for molecular data pertaining to the AIDS viruses, is one such important data resource, and the Los Alamos team led by Gerry Myers (co-PI) has played a key role in HIV sequence analysis. A flood of information is fast approaching about other pathogens and their continuing evolution; however, approaches to solving the complex problems of treatment of infectious diseases are seriously lagging behind the successful acquisition, processing and publishing of molecular data.

There exists a critical need for computational and theoretical efforts aimed at assimilating, integrating and interpreting vast collections of "atomized" biomedical information. This is the program we are developing at Los Alamos. The central task of the computation and theory effort is to synthesize the data into a coherent structure, and to construct, critique and interpret biochemical and immunological correlations in order to develop techniques for designing new drugs and therapies. The magnitude of the problems confronting health care nationally and internationally demands a highly
interdisciplinary scientific approach, as well as cooperative interactions between private industry and government.

3. Scientific Approach and Accomplishments

Following on our earlier work which analyzed correlated mutations in the V3 loop of HIV, we have developed an evolutionary model which incorporates non-independent mutations at different sequence positions in model sequences. This model takes into account the effects of the phylogenetic tree on the analysis of mutations, and shows a higher correspondence of correlated positions with structurally adjacent positions in the model than previous analyses.

Additional research under this project has dealt with analyzing the dynamics of HIV \textit{in vivo}. Using some simple models and patient data, we were able to establish for the first time some quantitative estimates of HIV dynamics \textit{in vivo}. We found that 10 billion viral particles are produced and cleared each day in a typical mid-stage AIDS patient. Further, we found that the vast majority of the virus (99\%) is produced by cells that only live about 1.5 days. These estimates suggest that if HIV replication can be completely stopped by potent drug combinations, then HIV and all cells infected by HIV may be able to be eliminated from the body after two to three years of completely effective treatment.

The kissing-loop RNA motif consists of two hairpins with complementary loop sequences. In retroviruses, one such motif is thought to facilitate the genome dimerization (Haddrick et al., 1996; Laughrea and Jette, 1996) and/or packaging (Berkhout and van Wamel, 1996). It is also a unique protein recognition site in regulation of ColEl plasmid DNA replication (Simon and Kleckner, 1988). In HIV, there exists a potential kissing-loop motif between a hairpin sequence in the TAR region and a homologous hairpin sequence (TAR*) in the GAG gene (about 800 bps downstream from the TAR sequence). We have found that the TAR-TAR* sequence motif is conserved in HIV-1, HIV-2, chimpanzee, as well as African green monkey viruses. Since the kissing-loop motif is thought to modulate different steps in the retroviral life cycle (e.g., recombination, translation, and encapsidation), it is very important to study its structure for the purpose of understanding its role in regulation as well as therapeutic functions.

Using an RNA folding protocol developed in our laboratory, we are studying the folding of this particular kissing-loop motif. The sequence and the secondary structure of the motif are shown in Fig. 1. Based on NMR imino proton spectra, Tinoco's lab has determined that five of the six bases in the loop region of the hairpin are involved in base-
pairings (Chang and Tinoco Jr., 1994). Using only the limited information of base-
pairings and the constraints based on the connectivities, we are modeling the structure and 
the flexibility of this particular structural motif. Initial structural models were constructed 
by short runs of Monte Carlo simulations using only the non-bonded energy. As a result, 
61 unique structures were generated. Each of these 61 initial modeled structures was then 
subjected to a refinement of 5,000 cycles of minimization using AMBER. The three 
predicted structures that have low conformational energies are shown in Fig. 2.

This past year we prepared two NIH grant proposals to fund research on human 
immunodeficiency virus (HIV) dynamics in vivo. The proposals were for collaborative 
research with David Ho, head of the Aaron Diamond AIDS Research Center (ADARC), 
Rockefeller University, New York. The first grant request was not funded but on 
resubmission it was ranked at the highest possible priority, first percentile of all proposals, 
and should be funded shortly as a subcontract from ADARC. A second NIH grant 
proposal, "Immune System Modeling" (A. Perelson, PI), has been funded beginning 
March 1996 for five years. Also in this past year we have prepared a grant proposal to 
DOE titled "Neural Net Approaches to Empirical Protein Potentials," which has been 
funded starting in FY 1997 for three years.

In other work we are laying the groundwork for a major new initiative to create a 
molecular database for diverse infectious pathogens. We have continuing discussion with 
the Centers for Disease Control concerning the establishment of an influenza database.

References

Berkhout, B. and van Wamel, J. L. "Role of the DIS hairpin in replication of human-

Chang K.-Y. and Tinoco Jr., I. "Characterization of a kissing hairpin complex derived 
from the human-immunodeficiency-virus genome," Proc. Natl. Acad. Sci. USA, 91: 

Haddrick, M., Lear, A. L., Cann, A. J. and Heaphy, S. J. "Evidence that a kissing loop 

Laughrea, M. and Jette, L. "HIV-1 genome dimerization: Formation kinetics and thermal-
stability of dimeric HIV-1(LAI) RNAs are not improved by the 1-232 and 296-790 regions 
flanking the kissing-loop domain," Biochem. 35: 9366-9374, 1996.

Simon, R. W. and Kleckner, N. "Biological regulation by antisense RNA in prokaryotes," 
Publications


Fig. 1. The sequence and the secondary structure of the HIV TAR-TAR* kissing-loop motif.
Fig. 2. Three predicted structures of the HIV TAR-TAR* kissing-loop motif. From the lightest to darkest shades are plotted a) helix 1, b) helix 2, c) helix 3, and d) the two single-stranded loops (p-A-p, p-U-p).