One of the enduring challenges facing HIV vaccine design is the remarkable rate of viral mutation and adaptation that limits the ability of the immune system to mount a lasting effective response. This rapid rate of mutation leads to extensive within- and between-host viral diversity that makes creation of a broadly reactive vaccine difficult. A first step in overcoming this challenge is to identify consistent patterns in viral adaptation. To this end, this dissertation proposes a new approach to modeling adaptation. This approach is applied to the largest available collection of HIV sequences, taken from chronically infected individuals for whom we also have detailed human genetic data. The result is a detailed prediction of how HIV interacts with, and adapts to, different immune systems, suggesting weak points in HIV's ability to adapt that may provide a window of opportunity for future vaccines.

Recent advances in DNA sequencing technology have vastly accelerated our ability to generate and test hypotheses about how biological systems interact. One popular approach is to compare the sequences of related species, looking for signature patterns of natural selection under the assumption that adaptation must be happening for a functional reason. If we can go one step further and predict the cause of the adaptation, then we are one significant step closer to understanding how a species interacts with its environment. Remarkably, HIV's rate of mutation is such that each patient is infected with a unique population of HIV that can be thought of as a distinct "quasi-species". Furthermore, the environment with which HIV interacts is primarily dominated by the human adaptive immune response. Thus, by understanding how HIV adapts to its environment, we will make progress in understanding how HIV thwarts our immune system, why some patients fare better than others, and, potentially, how we can give our immune systems the upper hand with a targeted vaccine.

Identifying patterns of adaptation can be thought of as a statistical exercise in identifying correlations. In essence, the goal is to identify HIV mutations that correlate with genetic variations in the human HLA proteins, a remarkably variable class of proteins that serve as the gateway to the human
immune response. The problem is that false correlations can be easily induced by the shared ancestry of HIV sequences: do infected patients in British Columbia harbor similar HIV mutations because they have similar human genetics, or because they are infected by similar HIV viruses that derived from a common ancestor? To answer this question, we can use evolutionary models that attempt to recapitulate the course of evolution and report the most likely scenario. The problem with such models is that they tend to be computationally and statistically inefficient and cannot scale to the complexity of HIV.

This dissertation proposes a new class of evolutionary models that can be used for inferring complex patterns of adaptation. The approach builds on the statistical machinery of dependency networks, a graphical representation in which nodes represent traits (HIV and/or human genetic variation) and arcs represent statistical dependencies among traits. By casting such networks in the context of evolutionary modeling, we are able to identify the complex interactions that characterize HIV adaptation.

In collaboration with four of the leading HIV labs around the world, we were able to pool two cohorts of HIV infected, HLA-typed individuals, creating the largest such cohort ever analyzed. Applying our model to these data revealed a startling complexity—yet an encouraging consistency—in the way HIV adapts to our immune systems. In addition to identifying specific mutations that allow HIV to evade the immune system, our model suggests that many such mutations trigger a cascade of other mutations that are required to stabilize the HIV protein. Such constraints suggest a crippling effect of some mutations that are (perhaps only partially) offset by other compensatory mutations, and which may provide weak points in the virus that are more vulnerable to immune attack.

By building on the machine learning approach of dependency networks, this dissertation proposes a new approach to evolutionary modeling. When applied to HIV, the result is a never-before-seen view into the intricacies of HIV adaptation—intricacies that may prove to be HIV’s Achilles’ heel.