

ABSTRACT

Noah G. Hoffman

**HIV-1 Sequence Variability Reveals Determinants of Viral Phenotype.
(Under the direction of Dr. Ronald Swanstrom.)**

Exposed on the surface of the human immunodeficiency virus type I (HIV-1), gp120 is a target of both the humoral and cellular immune responses, and is the most variable viral protein. Escape from the immune system, however, cannot come at the expense of gp120 function, which includes the ability to bind to both CD4 and a “coreceptor” (CCR5, CXCR4, or both). Distinguishing the sequence and structural determinants of receptor binding and coreceptor preference from the “noise” of gp120 variability is an important goal of HIV-1 research.

In this thesis, I describe bioinformatic approaches to revealing functional determinants in HIV-1 gene products through studies of sequence variability. Chapter 2 tests the reliability of algorithms for inferring HIV-1 entry phenotype using sequences from the third variable loop (V3) of gp120. In Chapter 3, we used one of these phenotyping methods to classify HIV-1 subtype B gp120 sequences as X4-like (CXCR4-using) or R5-like (CCR5-using). Using these classifications, we identified positions outside of V3 where either amino acid composition or variability was linked to inferred phenotype. This study both confirms previous observations, and predicts specific positions that contribute to a functional relationship between the V2, V3, and C4 regions of gp120.

Group M HIV-1 isolates are classified into phylogenetically-defined clusters called *subtypes*. Although subtype B X4 variants are common, subtype C X4 isolates rarely emerge. In Chapter 4, we propose that this difference is due to subtype-specific structural features of

V3. Accordingly, we demonstrate that patterns of amino acid variability, composition, and covariation in subtype B and C V3 sequences are consistent with distinct V3 conformations. In addition, we report subtype-specific differences in the binding of an antibody recognizing a conformational epitope in V3.

Finally, we have identified linked sequence changes in HIV-1 subtype B protease using translated sequences from protease inhibitor-treated and -untreated subjects (Chapter 5). Both positive and negative associations between positions in protease were detected. Structural proximity suggests that numerous pairs may interact within a local environment.