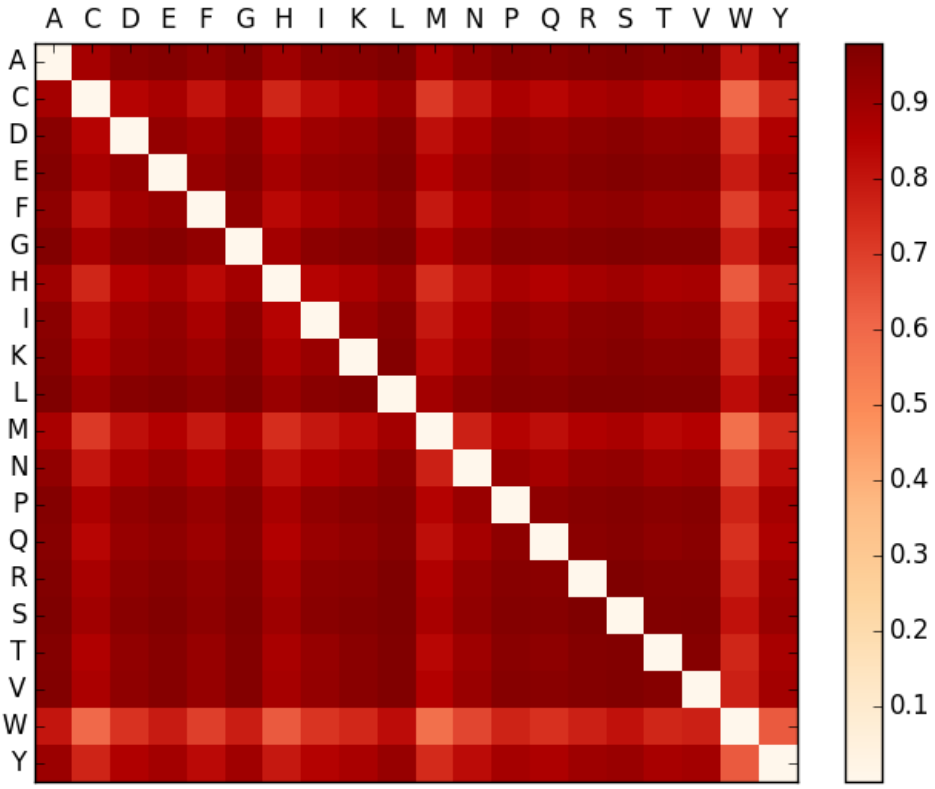

Figure 1:

1 Results of the Simultions - Idealized Model

1.1 An upper bound on proteins uniquely identifiable by two-color barcoding

Each protein in a given proteome is fingerprinted by a string of length l, where l is a two-color reduction of the parent sequence. For example, for an "EY" fingerprint of the sequence MYTARGETPRQTEIN would be YEE. The number of fingerprints with a unique parent protein, N_f, is compared to the total number of proteins in the proteome, N_p. We define uniqueness in the idealized model as $\frac{N_f}{N_p}$. Here, we examine all 20^2 possible two-color fingerprints. Discarding the diagonal, we find that a 'LS' fingerprint maximizes the number of uniquely identifiable human proteins (97.9%) while a 'MW' fingerprint minimizes the number of uniquely identifiable human proteins (57.9%). If we consider only fingerprints with tractable synthetic strategies , such as a 'KC' fingerprint, we find that 86.4% of human proteins are in principle uniquely identifiable.