Supporting Text

Methods

These were standardized to be unitless with mean zero and variance one. Let A_{ij} denote the amino acid from sequence i in position j of the alignment, let s be any standardized scale, and let $s(A_{ij})$ be the value of amino acid A_{ij} on that scale. For each position j in the alignment, the positional mean and variance were computed, respectively, as $\mu_j = \sum_i w_i s(A_{ij})$ and $\sigma_j^2 = \sum_i w_i \left(s(A_{ij}) - \mu_j\right)^2$. We supposed that the s-specific disagreement between the alignment column and an amino acid z with value $s(z)$ could be quantified by $\frac{s(z) - \mu_j}{\sigma_j}$. By folding together the specific disagreement values for each property, an aggregate score (called a MAPP score) was developed to test the overall incompatibility of amino acid z to position j. We compared MAPP scores to an F-distribution to assess the impact of candidate substitutions.