



Figure S6: Example of matrix F containing the frequencies of the 136 possible states for positions NT-4 and NT-5 in both semisequences. Due to the equivalence between complementary combinations (see Text S1), semisequences 1 and 2 do not necessarily correspond to the left and right ones, respectively. This matrix is that associated to the recognition amino acids $K_{15}S_{16}$ (BS-logo in insert, all logos in Appendix of Text S1). In red those statistically significant frequencies. Columns and rows corresponding to the semisequences involved in at least one significant state are highlighted in soft gray. Crossing cells (in dark gray) constitute the triangular matrix S (see Text S1 for details).