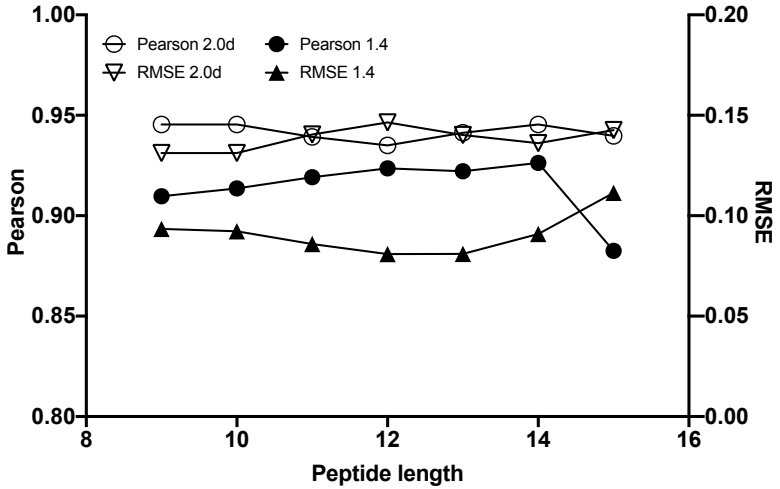
Suppl. Figure 1

Scanning through different peptide lengths - ANN training (2.0d and 1.4)



Optimization of peptide length. Peptides of lengths 9-15 were synthesised on a peptide array and incubated with HLA-DRB1*03:01. Data from each peptide length were used to train ANNs and the internal correlation and RMSE output from the networks were used to find the optimal peptide length. NNAlign_2.0d (open) shows little or no response to the peptide length input in contrast to NNAlign_1.4 (solid) which clearly show optimal performance at peptide length = 13 (Pearson to RMSE). The decision to optimise the peptide length rests on two theoretical concerns. The primary concern is that of quality of synthetic peptides vs. their length. Each elongation step of a solid-phase peptide synthesis carries a certain risk of failure, and the longer the peptide, the greater the risk of one or more failures having occurred. Thus, everything else being equal, an approach using shorter peptides should generate higher quality data than an approach using longer peptides. When that is said, one should not use a length shorter than what is needed for HLA-DR binding. The core region of a peptide binding to HLA-DR is thought to be around 9 amino acids (as it is for HLA class I). In some cases, a 9mer sequence will suffice; however, shorter elongations at either end tend to improve binding. Since such longer versions are likely to be generated during natural processing, selecting 9mers as the length examined would risk missing longer versions naturally processed, presented and recognized. A secondary concern is that of “stressing” the pattern recognition ability of the neural networks by increasing the diversity of the sequence space that needs to be considered. Everything thing else being equal, increasing the number of subsequences that potentially could be generated from longer peptides (and considering the increased risk of synthesis errors, which would generate erroneous results) would increase the number of possible solutions (and errors) that can be assigned to each peptide, and increase the inherent “noise” in the data. We systematically examined different lengths experimentally and identified 13mers as the optimal length.