The following data files are available associated with Bisarad et al, “**Predicting Monoclonal Antibody Binding Sequences from a Sparse Sampling of All Possible Sequences”:**

|  |  |
| --- | --- |
| **File** | **Description** |
| Averaged\_Array\_Data.xlsx | This is the average array binding data from 4 repetitions. It contains the peptide sequences, a binary flag that can be used to remove the cognate sequences of the monoclonal antibodies used in binding from the dataset and 32 columns of binding data (4 concentrations, 8 mAbs). The first row contains the mAb names and the second row contains the mAb concentrations. All subsequent rows show the binding of the mAb at that concentration to each sequence, measured in fluorescent counts. Note that the photon detector saturates at 2^16 counts. |
| Dataset\_rep1.xlsx | Same format as above, but the unaveraged data for repetition 1 |
| Dataset\_rep2.xlsx | Same format as above, but the unaveraged data for repetition 2 |
| Dataset\_rep3.xlsx | Same format as above, but the unaveraged data for repetition 3 |
| Dataset\_rep4.xlsx | Same format as above, but the unaveraged data for repetition 4 |
| Figure2\_data.xlsx | The measured and predicted data used to make the matrix of Figure 2 |
| Figure3\_data.xlsx | The sequence and predicted binding as a function of sequence to the proteins in Figure 3 |

In addition to the data, there is also a Matlab script, “mAb\_sequence\_binding\_relationship\_updated.m” which will perform the neural network analysis for any mAb in the dataset and create a model. It uses that model to predict the binding of the known mAb epitope as well as 1 million randomly determined sequences and then provides the rank of the mAb in that sequence set. It will run on any of the five datasets provided. Matlab 2022a was used to create this.