Enhancing Docking Accuracy with PECAN2, a 3D Atomic Neural Network Trained without Co-Complex Crystal Structures

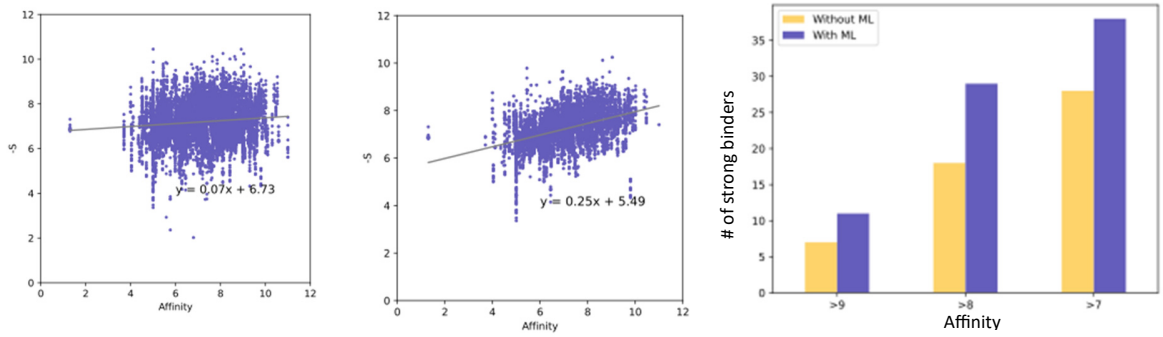
Heesung Shim [1,*], Jonathan E. Allen [2] and W. F. Drew Bennett [1,*]

[1] Physical and Life Sciences, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA
[2] Global Security Computing Applications Division, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA; allen99@llnl.gov
* Correspondence: shim2@llnl.gov (H.S.); bennett69@llnl.gov (W.F.D.B.)

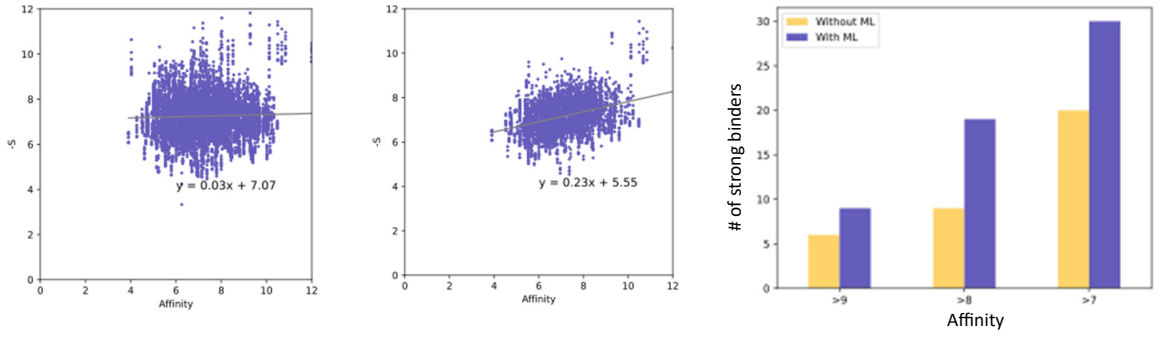
Figure S1 displays a correlation scatter plot between docking scores and experimental ligand affinities for all docking poses in the random split experiment (Section 3.1.1). It also shows the count of ligands within the top 5% ranked compounds with affinities of 7, 8, and 9 or higher. Upon the application of the PECAN2 pose classifier, a notable improvement in the Pearson correlation coefficients between binding affinity and docking score was observed, in contrast to scenarios

where PECAN2 was not utilized. Furthermore, an examination of compounds within the top 5% revealed an increase in the count of strong binders.

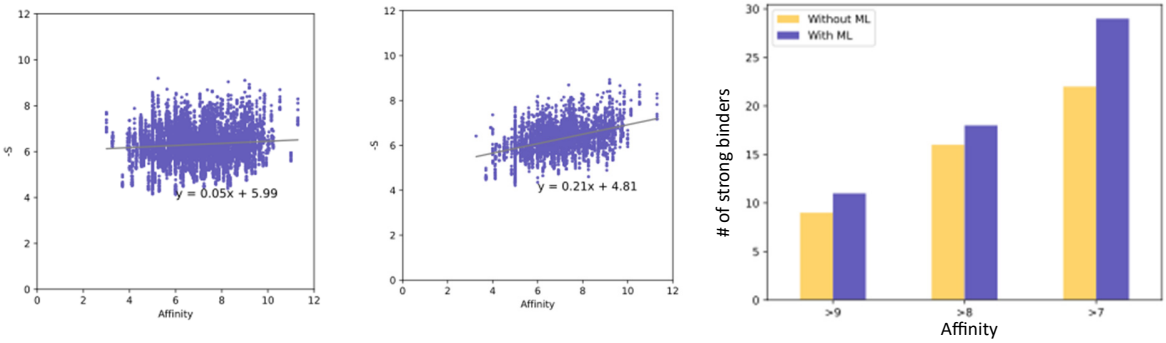
(a) Mu—MOE



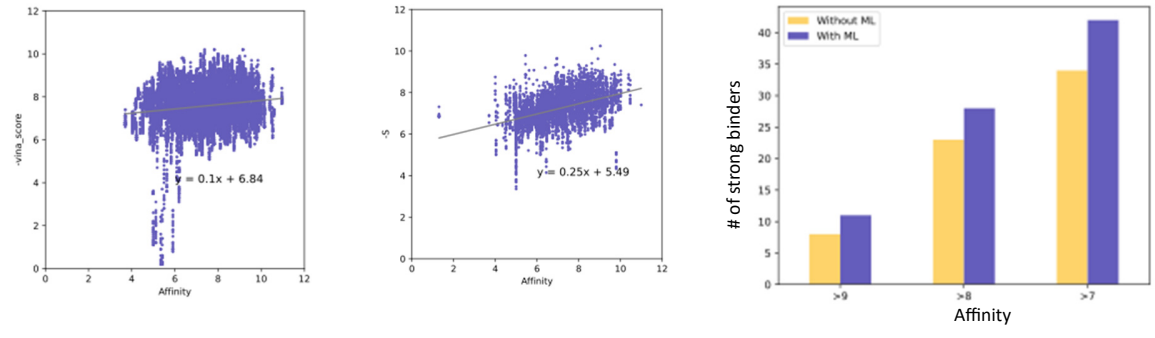
(b) Kappa—MOE



(c) Delta—MOE



(d) Mu—Autodock Vina



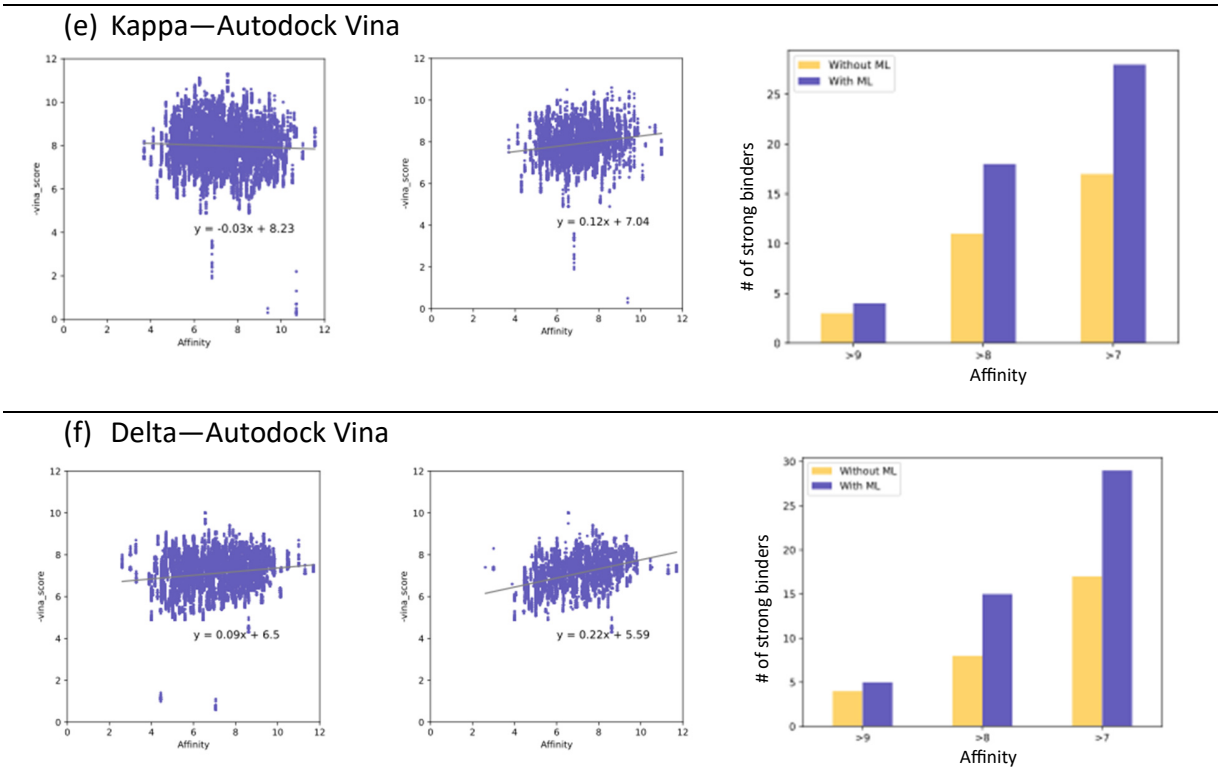
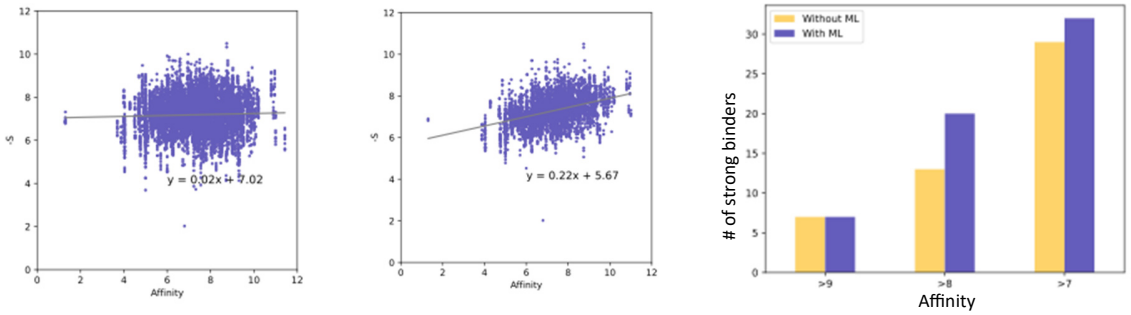


Figure S1: Random split (Section 3.1.1)—Scatter plot between docking scores and experimental ligand affinities for all docking poses (left) and a histogram indicating the count of strong binders (right). (a, b, c) The scatter plots on the left represent the results using only docking with the MOE program for mu opioid data (a), kappa opioid data (b), and delta opioid data (c) without PECAN2, while the ones on the right show the poses remaining after filtering out uncorrelated poses using PECAN2. In the histograms on the right, the yellow color represents the scenario without using PECAN2 with mu opioid data (a), kappa opioid data (b), and delta opioid data (c) using only docking with the MOE program, while the purple color represents the scenario with the use of PECAN2. (d, e, f) The scatter plots on the left represent the results using only docking with the Autodock Vina program for mu opioid data (d), kappa opioid data (e), and delta opioid data (f) without PECAN2, while the ones on the right show the poses remaining after filtering out uncorrelated poses using PECAN2. In the histograms on the right, the yellow color represents the scenario without using PECAN2 with mu opioids data (d), kappa opioid data (e), and delta opioid data (f) using only docking with the Autodock Vina program, while the purple color represents the scenario with the use of PECAN2.

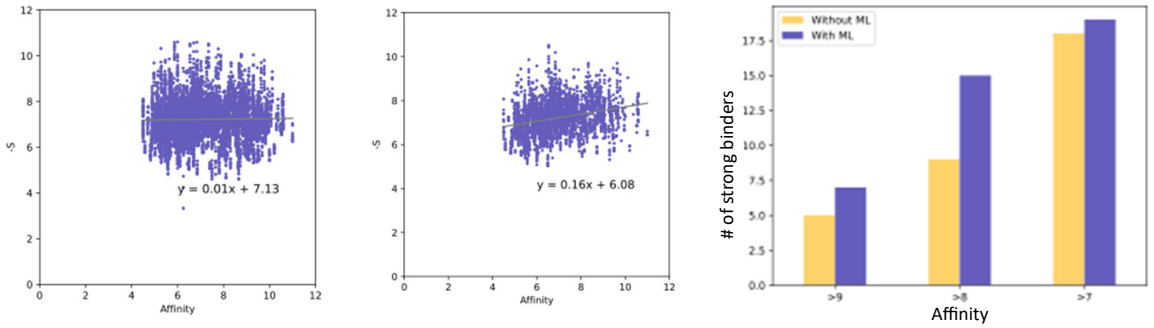
Figure S2 displays a correlation scatter plot between docking scores and experimental ligand affinities for all docking poses in the scaffold split experiment (Section 3.1.2). It also shows the

count of ligands within the top 5% ranked compounds with affinities of 7, 8, and 9 or higher. In general, akin to the random split data experiment, the use of the PECAN2 pose classifier led to enhanced Pearson coefficients and the identification of more strong binders compared to scenarios without PECAN2. However, as anticipated, the overall improvement was lower than that observed in the random split.

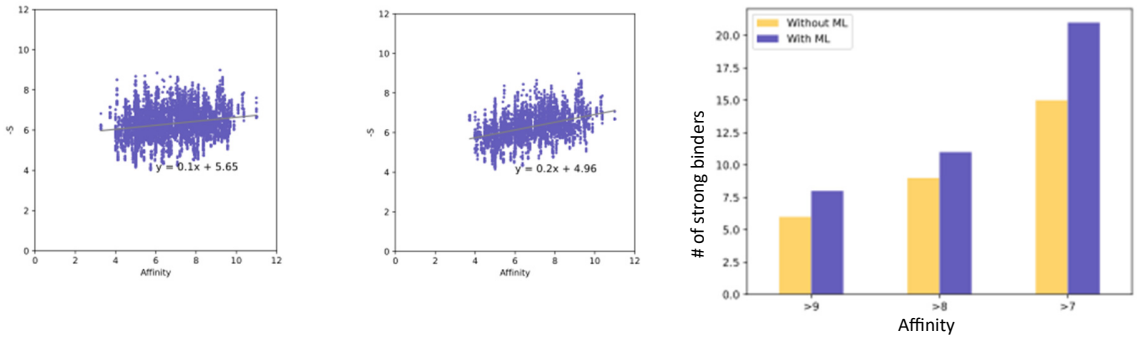
(a) Mu—MOE



(b) Kappa—MOE



(c) Delta—MOE



(d) Mu—Autodock Vina

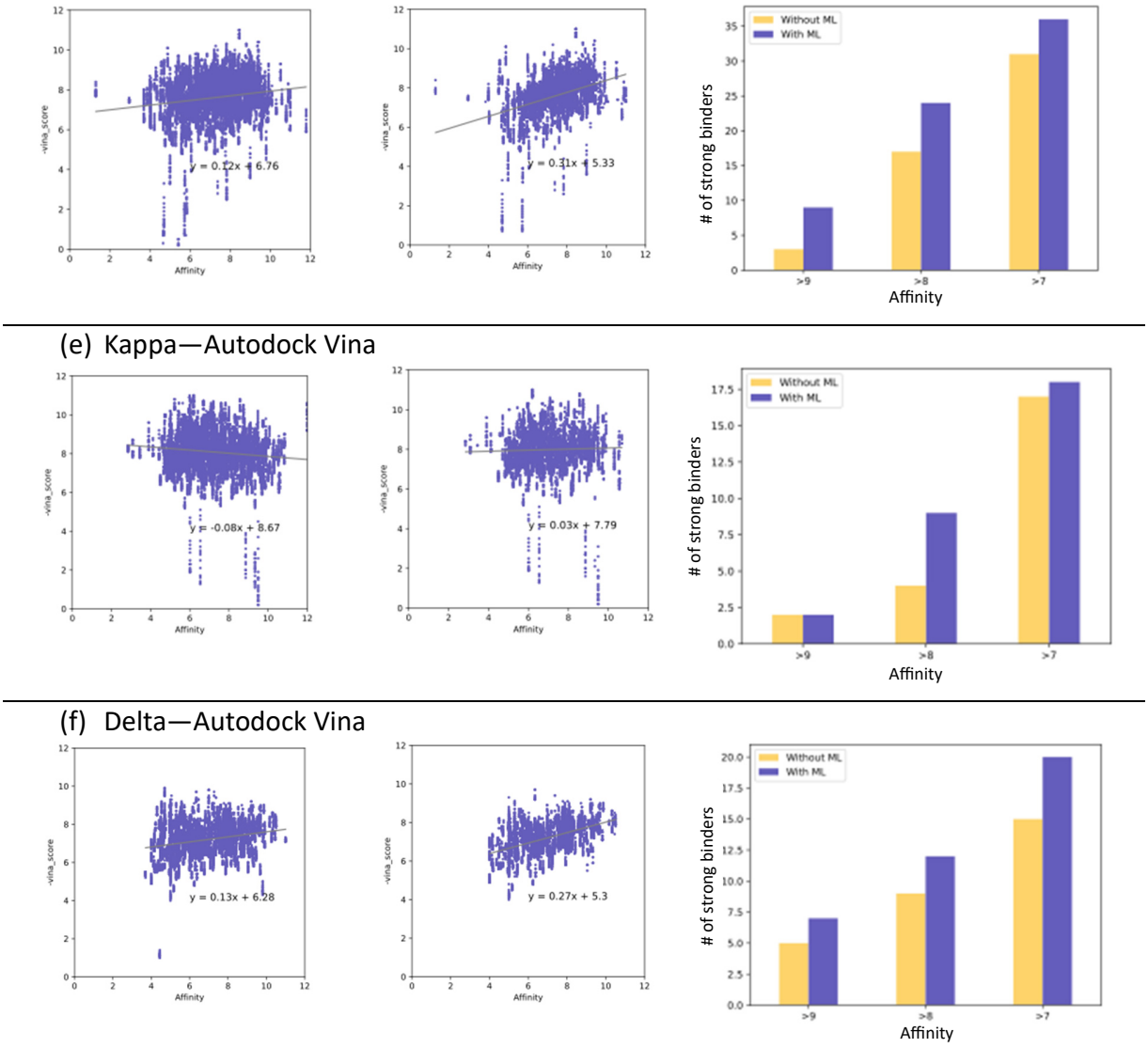


Figure S2: Scaffold split (Section 3.1.2)—Scatter plot between docking scores and experimental ligand affinities for all docking poses (left) and a histogram indicating the count of strong binders (right). (a, b, c) The scatter plots on the left represent the results using only docking with the MOE program for mu opioid data (a), kappa opioid data (b), and delta opioid data (c) without PECAN2, while the ones on the right show the poses remaining after filtering out uncorrelated poses using PECAN2. In the histograms on the right, the yellow color represents the scenario without using PECAN2 with mu opioid data (a), kappa opioid data (b), and delta opioid data (c) using only docking with the MOE program, while the purple color represents the scenario with the use of PECAN2. (d, e, f) The scatter plots on the left represent the results using only docking with the Autodock Vina program for mu opioid data (d), kappa opioid data (e), and delta opioid data (f) without PECAN2, while the ones on the right show the poses remaining after filtering out uncorrelated poses using PECAN2. In the histograms on the right, the yellow color represents the scenario without using PECAN2 with mu opioid data (d), kappa opioid data (e), and delta opioid

data (f) using only docking with the Autodock Vina program, while the purple color represents the scenario with the use of PECAN2.

Figure S3 depicts the results of testing PECAN2 on the SARS2 Mpro data after training it with the delta opioid receptor docking data. As anticipated, this model not only failed to improve the Pearson correlation in the SARS2 Mpro data but also resulted in a decrease. The Pearson value dropped from 0.261 to -0.006, and PECAN2 trained with delta data identified even fewer strong binders than when PECAN2 was used alone. This is in line with our expectations due to the significant differences in protein structures between the two datasets.

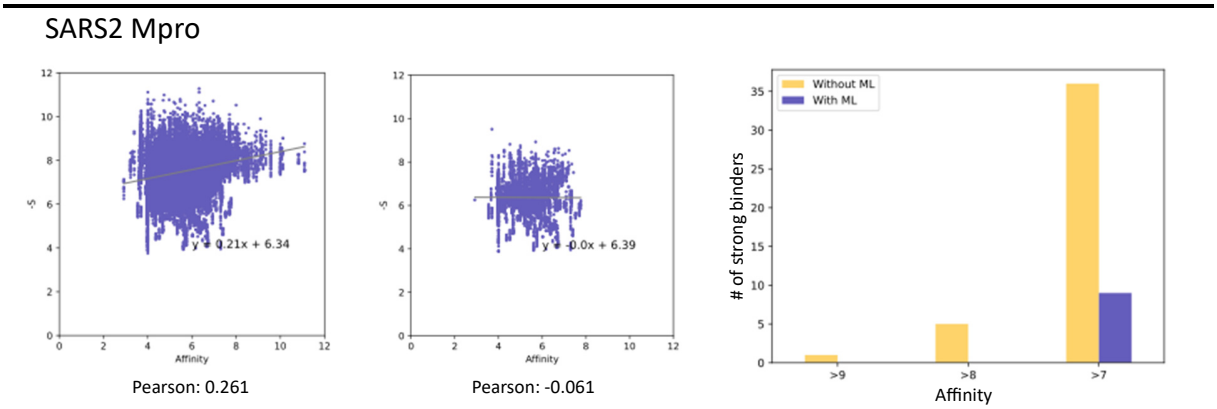


Figure S3: Scatter plot between MOE docking scores and experimental ligand affinities for all docking poses of SARS2 Mpro data (left and middle) and a histogram indicating the count of strong binders (right). The scatter plot on the left represents the results using only docking without PECAN2, while the one in the middle shows the poses remaining after filtering out uncorrelated poses using PECAN2. In the histogram on the right, the yellow color represents the scenario without using PECAN2, while the purple color represents the scenario with the use of PECAN2.

Figure S4 displays a correlation scatter plot between docking scores and experimental ligand affinities for all docking poses of the SARS2 Mpro experiment (Section 3.1.5). It also shows the count of ligands within the top 5% ranked compounds with affinities of 7, 8, and 9 or higher. In general, similar to the opioid receptor experiment, the use of the PECAN2 pose classifier led to enhanced Pearson coefficients and the identification of more strong binders compared to scenarios without PECAN2.

(a) MOE

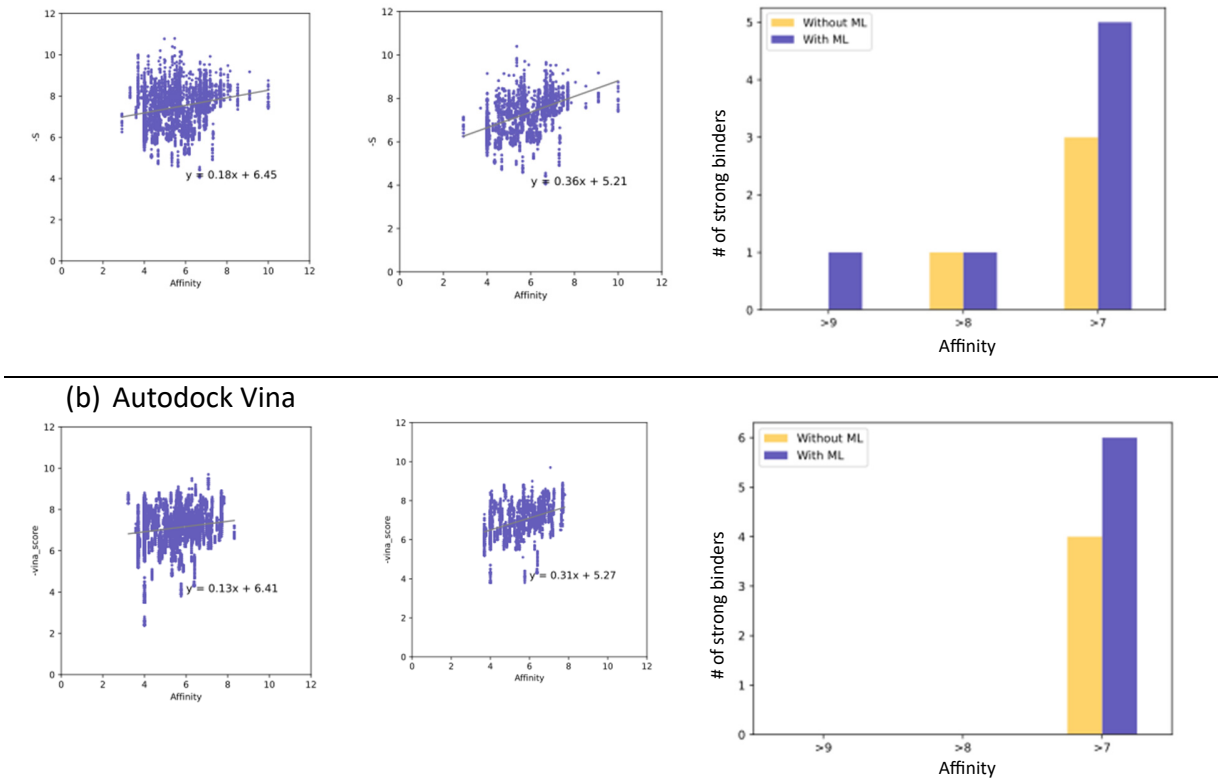


Figure S4: Scatter plot between docking scores and experimental ligand affinities for all docking poses of SARS2 Mpro data (left and middle) and a histogram indicating the count of strong binders (right). (a) The scatter plots on the left represent the results using only docking with the MOE program for SARS2 Mpro data without PECAN2, while the ones on the right show the poses remaining after filtering out uncorrelated poses using PECAN2. In the histograms on the right, the yellow color represents the scenario without using PECAN2 with for SARS2 Mpro data using only docking with the MOE program, while the purple color represents the scenario with the use of PECAN2. (b) The scatter plots on the left represent the results using only docking with the Autodock Vina program for SARS2 Mpro data without PECAN2, while the ones on the right show the poses remaining after filtering out uncorrelated poses using PECAN2. In the histograms on the right, the yellow color represents the scenario without using PECAN2 with for SARS2 Mpro data using only docking with the Autodock Vina program, while the purple color represents the scenario with the use of PECAN2.