Team name	Northeastern University Warriors of the Anti-Viral Enterprise (NUWAVE)
Team member(s) (firstname lastname; …)	Mary Jo Ondrechen (PI), **Suhasini Iyengar**, Kelton Barnsley, Yen Vu, Penny Beuning, Ian Bongalonta, Alyssa Herrod, Jasmine Scott
Affiliation	Northeastern University, USA
Contact email	Iyengar.s@northeastern.edu ; mjo@neu.edu
Contact phone number (optional)	+1-857-415-9653 (Iyengar) +1-508-740-9513 (Ondrechen)
Protein targets (for example: 3CLPro/Nsp5, BoAT1, Fc Receptor, Furin, IL6R, M protein, Nspx, OrfXx, N, E, etc…) | 3 required	**NSP1**, RNA Methyltransferase (NSP16-NSP10), N Protein, Main Protease (monomer), Main Protease (Dimer), NSP3 (this report talks about the **NSP1**)

Section 1: Methods:

Part A- Binding site prediction:

For the binding site prediction, Partial Order Optimum Likelihood (POOL) (1, 2) was used. Partial Order Optimum Likelihood (POOL) is a machine learning method that predicts biochemically active sites using the three-dimensional structure of the query protein as input. POOL predicts multiple types of binding sites in proteins which include catalytic sites, allosteric sites and other sites, some of which may not be detected by other predictive methods. POOL generates a rank ordered list of all the amino acids in the protein structure in the order of likelihood of biochemical activity. POOL predicts some sites that might be overlooked by other methods because POOL is based primarily on computed electrostatic and chemical properties (3,4) of the query protein, rather than a purely informatics-based approach. POOL points to the residues involved in reversible binding, including catalytic sites, non-catalytic binding sites such as allosteric sites, ligand transport sites, and some protein-protein interaction sites. The other input features for POOL consist of properties of the local environment (1,2) and surface topological metrics (5).

Part B- Molecular Docking:

Molecular Docking was performed using Schrödinger Glide (6). For docking in Schrödinger Glide, the ligands were prepared using LigPrep (7), the protein was minimized and optimized using Protein Preparation Wizard and the grid for docking was prepared using Receptor Grid Generation using the top 10 % of the POOL predicted residues as the centroid for ligand placement in Schrödinger 2019-3. Molecular Docking was performed on the Discovery Cluster at the Massachusetts Green High-Performance Computing Center using Glide. Glide Standard Precision (SP)(8) was used a filter to remove false positive results and top predicted ligands with docking score of <=-7 kcal/mol were used for Glide Extra Precision (XP) (9).

Section 2: Targets

Target 1: NSP1

A comparative model structure was built using the homology modeling feature in the YASARA (10) suite of programs. Only one viable template structure was available, the 2007 NMR structure of the SARS coronavirus NSP1 (11). Molecular dynamics simulations on the model structure were performed in YASARA (10) in a water box with a 10 Å margin around

all atoms and with the YAMBER3 force field. Simulations were performed at a temperature of 310K, pH 7.0, a solvent density of 0.993 g/cc and an NaCl concentration of 0.9%. A time step of 1.0 fs was used. Following equilibration, dynamic structures were targeted for ensemble docking, utilizing pockets defined by residues predicted by POOL [1,2] for each dynamic structure. Before running POOL on this structure, it was analyzed on YASARA and pKa prediction and energy minimization using YAMBER3 force field were done for this model. This structure was further prepared before docking using the Protein Preparation Wizard on Maestro. The protein preparation wizard allows the user to take the protein in its raw state-which might be missing hydrogen atom and have incorrect bond orders-and convert it into a state which is properly prepared for use by Schrodinger products such as Glide. Protein Preparation step on Maestro contains three basic steps- first is preprocessing the protein structure. This step performs the basic calculations for assigning bond orders, adding hydrogens, creating disulfide bonds, filling missing side chain or missing loops, deleting waters among many others whenever needed. The second step is protein refinement. This step consists of optimization of the hydrogen bond network by reorienting the hydroxyl and thiol groups, water molecules, amide groups of asparagine (Asn) and glutamine (Gln), and the imidazole ring in histidine (His); and predicting protonation states of histidine, aspartic acid (Asp) and glutamic acid (Glu) and tautomeric states of histidine. The last step is Restrained minimization which provides controls for optimizing the corrected structure, to relieve any strain and fine-tune the placement of various groups. NOTE: The NSP1 model structure is also attached with the files in the .zip folder.

Section 3: Libraries

The ligands were obtained from the following databases:

a) ZINC FDA library (https://zinc15.docking.org/substances/subsets/fda/)
b) CAS Antiviral set (https://www.cas.org/covid-19-antiviral-compounds-dataset)
c) Enamine FDA library (https://enamine.net/hit-finding/compound-collections/bioreference-compounds/fda-approved-drugs-collection)
d) Antiviral library consisting of compounds from- Selleck Chemicals Antiviral Library Enamine Antiviral Library and Asinex Antiviral Library

The ligands from all these libraries were prepared using LigPrep tool in Maestro. Ligprep is a tool designed to prepare high quality all-atom 3D structures for large numbers of drug-like molecules. The LigPrep process consists of a series of steps that perform conversions, apply corrections to the structures, generate variations in the structure, eliminate unwanted structures and optimize all the structures.

Section 4: Results

Part A: POOL- Prediction of Binding Sites:

POOL generates a rank-ordered list of all the amino acids in a protein structure, in order of likelihood of biochemical activity. The POOL predicted sites for the NSP1 protein are as follows.

2GLU, 3SER, 4LEU, 13HIS, 45HIS, 46LEU, 50THR, 51CYS, 87GLU, 109PRO, 110HIS, 113GLU, 134HIS, 137GLY.

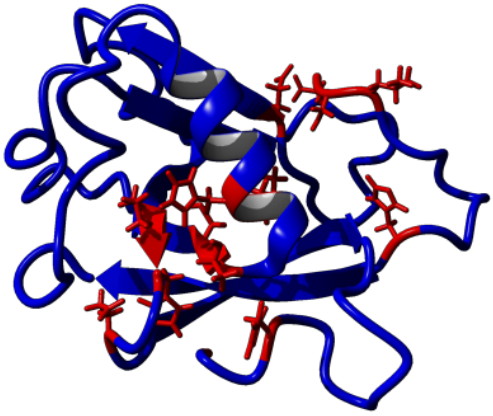

Figure 1: POOL Predicted residues shown in Red Blue for NSP1 structure at 300 ns.

Section B: Molecular Docking:

From the libraries using for testing the top hits were obtained from the CAS Antiviral library. Glide SP docking was performed on the entire library and the top hits from Glide SP were given as input to Glide XP. The results tabulated below are from Glide XP.

cas.rn	docking score	XP GScore	Interactions **BOLD-POOL** H-bond, PI-PI, salt bridge, halogen bond, Pi-cation
1002334-89-5	-11.614	-11.614	Val86, Leu122, Arg124, Asn126, Leu46, Lys47
752233-03-7	-11.091	-11.121	Asp33,Asp33 Ser34, Glu37, glu37, Glu41, Glu41, Gln44, Lys47, Arg124, Arg124, Val86, Leu88, His13
1002334-95-3	-11.016	-11.016	Arg124, Asn126,Lys47, Leu46, Val86, **Leu4**
36078-14-5	-10.834	-11.027	Lys47, Arg124, Arg124,**Leu4**, Leu88, Val86
926902-29-6	-10.756	-10.760	Arg124, Arg124, Lys125, Lys47, Lys47, Arg443, **Leu4**

References:

1. Tong, W., Y. Wei, L.F. Murga, M.J. Ondrechen, and R.J. Williams, Partial Order Optimum Likelihood (POOL): Maximum likelihood prediction of protein active site residues using 3D structure and sequence properties. PLoS Comp Biol, 2009. 5(1): e1000266.
2. Somarowthu, S., H. Yang, D.G. Hildebrand, and M.J. Ondrechen, High-performance prediction of functional residues in proteins with machine learning and computed input features. Biopolymers, 2011. 95(6): 390-400.

3. Ondrechen, M.J., J.G. Clifton, and D. Ringe, THEMATICS: A simple computational predictor of enzyme function from structure. Proc. Natl. Acad. Sci. (USA), 2001. 98: 12473-12478.
4. Ko, J., L.F. Murga, P. André, H. Yang, M.J. Ondrechen, R.J. Williams, A. Agunwamba, and D.E. Budil, Statistical criteria for the identification of protein active sites using theoretical microscopic titration curves. Proteins, 2005. 59(2): 183-195.
5. Capra, J.A., R.A. Laskowski, J.M. Thornton, M. Singh, and T.A. Funkhouser, Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. PLoS Comput Biol, 2009. 5(12): e1000585.
6. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shaw, D. E.; Shelley, M.; Perry, J. K.; Francis, P.; Shenkin, P. S., "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy," J. Med. Chem., 2004, 47, 1739–1749.
7. Schrödinger Release 2019-3: Glide, Schrödinger, LLC, New York, NY, 2019.
8. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L., "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening," J. Med. Chem., 2004, 47, 1750–1759
9. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren,T. A.; Sanschagrin, P. C.; Mainz, D. T., "Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes," J. Med. Chem., 2006, 49, 6177–6196.
10. Krieger, E., K. Joo, J. Lee, J. Lee, S. Raman, J. Thompson, M. Tyka, D. Baker, and K. Karplus, Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. Proteins 2009. 77 Suppl 9: 114-122.
11. Almeida, M.S., M.A. Johnson, T. Herrmann, M. Geralt, and K. Wüthrich, Novel beta-barrel fold in the nuclear magnetic resonance structure of the replicase nonstructural protein 1 from the severe acute respiratory syndrome coronavirus. Journal of virology, 2007. 81(7): 3151-3161