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 **NSP3**)

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: NSP3

Three structures of SARS-CoV-2 NSP3 were targeted for docking, 6W02 (10), 6WCF (11), and 6WEY (12), utilizing pockets defined by residues predicted by POOL [1,2] for each structure. The structures were 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 Schrödinger 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.

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

4A: 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 top POOL predicted residues for the NSP3 protein structures are :

6W02 – His91, Leu123, Asn37, Lys28
6WCF – His45, Cys92, Ala38, Cys81, Asn37, Ala50, Val49, Pro125, Tyr17, Ala39, Lys11, Tyr152

6WEY – His249, Cys296, His298, Val253, Ala242, Tyr317, Cys285, Ala254, Asn241, Pro329, Gly252, Cys347, Ala293 (Note different sequence numbering scheme)

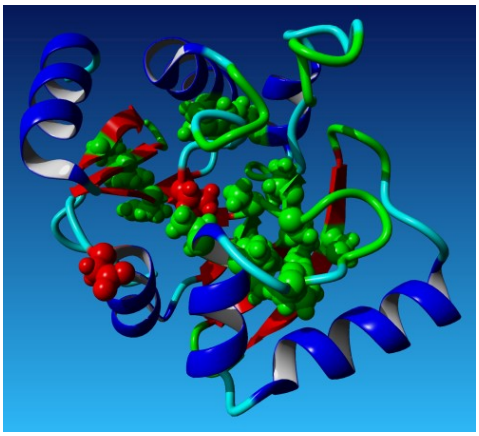

Figure 1. POOL-predicted residues, shown as ball-and-stick, in the NSP3 structure 6WCF (11).

4B: Molecular Docking:

Glide SP docking was performed on the libraries and the top hits from Glide SP were given as input to Glide XP. The results tabulated below are from Glide XP.

Table 1. Top Hits for NSP3 Structures

Compound ID	XP score	Target-PDBID	Ligand Library	Compound ID
ZINC000002005305	-12.108	nsp3-6W02	ZINC FDA	ZINC000002005305
ZINC000001536109	-11.201	nsp3-6W02	ZINC FDA	ZINC000001536109
1988747-23-4	-11.132	nsp3-6WCF	CAS	1988747-23-4
ZINC000002005305	-10.666	nsp3-6W02	ZINC FDA	ZINC000002005305
ZINC000002005305	-10.666	nsp3-6W02	ZINC FDA	ZINC000002005305
ZINC000003806413	-10.638	nsp3-6WCF	ZINC FDA	ZINC000003806413
ZINC000009212427	-10.442	nsp3-6WCF	ZINC FDA	ZINC000009212427
ZINC000008214418	-10.43	nsp3-6WCF	ZINC FDA	ZINC000008214418
ZINC000085540215	-10.43	nsp3-6WCF	ZINC FDA	ZINC000085540215
ZINC000002005305	-10.309	nsp3-6WCF	ZINC FDA	ZINC000002005305
ZINC000003813010	-10.169	nsp3-6WCF	ZINC FDA	ZINC000003813010
ZINC000085540219	-10.138	nsp3-6WCF	ZINC FDA	ZINC000085540219
ZINC000085540223	-10.138	nsp3-6WCF	ZINC FDA	ZINC000085540223
ZINC000003842753	-9.775	nsp3-6W02	ZINC FDA	ZINC000003842753
ZINC000003861768	-9.775	nsp3-6WCF	ZINC FDA	ZINC000003861768
2088572-78-3	-9.761	nsp3-6WCF	CAS	2088572-78-3
ZINC000003843198	-9.738	nsp3-6WCF	ZINC FDA	ZINC000003843198
ZINC000003842753	-9.711	nsp3-6W02	ZINC FDA	ZINC000003842753
1977515-69-7	-9.319	nsp3-6WCF	CAS	1977515-69-7
ZINC000000137884	-9.319	nsp3-6W02	ZINC FDA	ZINC000000137884
ZINC000000156792	-9.291	nsp3-6WCF	ZINC FDA	ZINC000000156792
ZINC000000968255	-9.291	nsp3-6WCF	ZINC FDA	ZINC000000968255
ZINC000003842753	-9.241	nsp3-6W02	ZINC FDA	ZINC000003842753
ZINC000003843198	-9.217	nsp3-6W02	ZINC FDA	ZINC000003843198
ZINC000003813010	-9.186	nsp3-6W02	ZINC FDA	ZINC000003813010
2093559-18-1	-9.099	nsp3-6WCF	CAS	2093559-18-1

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. Michalska, K., Kim, Y., Jedrzejczak, R., Maltseva, N., Endres, M., Mececar, A., Joachimiak, A., "Crystal Structure of ADP ribose phosphatase of NSP3 from SARS CoV-2 in the complex with ADP ribose," PDB ID 6W02, 2020, to be published.
11. Michalska, K., Kim, Y., Jedrzejczak, R., Maltseva, N., Endres, M., Mececar, A., Joachimiak, A., "Crystal Structure of ADP ribose phosphatase of NSP3 from SARS CoV-2 in the complex with MES," PDB ID 6WCF, 2020, to be published.
12. Vuksanovic, N., Melkonian, T.R., Serrano, D.A., Virdi, R.S., Dahal, N., Frick, D.N., Silvaggi, N.R., "High-resolution structure of the SARS-CoV-2 NSP3 Macro X domain," PDB ID 6WEY, 2020, to be published.