Impactful Mutations in Mpro of the SARS-CoV-2 Proteome

Gideon Wolfe
Western Washington U.
Bellingham, WA, USA

Othmane Belhoussine
Western Washington U.
Bellingham, WA, USA

Anais Dawson
Western Washington U.
Bellingham, WA, USA

Filip Jagodzinski
Western Washington U.
Bellingham, WA, USA

ABSTRACT

We explore how amino acid mutations affect the stability of a 306 residue main protease of the COVID-19 proteome. We employ two computational approaches, Site Directed Mutagenesis (SDM) and short runs of Molecular Dynamics. We focus our attention on residues 25-32 that make up a beta sheet of a canonical beta barrel close to an active site which includes Histidine 41. We considered this region a good candidate for mutations because such a large perturbation of a highly structured region close to the active site may prove to be highly detrimental to the protein's stability and may affect catalytic efficiency. We mutated the 8 residues *in silico* to all other possible amino acids, and analyzed the resulting 152 mutants. Both computational methods predict that only a few specific mutations to some of the 8 residues have a major effect on the structural stability of the protein.

CCS CONCEPTS

• **Applied computing → Molecular structural biology**.

KEYWORDS

SARS-CoV-2, mutations, computational biology

ACM Reference Format:
Gideon Wolfe, Othmane Belhoussine, Anais Dawson, and Filip Jagodzinski. 2018. Impactful Mutations in Mpro of the SARS-CoV-2 Proteome. In *Proceedings of ACM-BCB '20: Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB '20).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Since December 2019, a new respiratory disease, caused by a member of the Coronaviridae family, has been responsible for the current global pandemic which is termed coronavirus disease 2019 (COVID-19) [2][11]. The virus responsible has since been sequenced and named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The World Health Organization reports that millions have been infected, and hundreds of thousands have died worldwide[6].

The virus genome has been sequenced, which contains more than 100 proteins, many of which are the focus of intense research efforts that aim to develop drugs to combat the pandemic. The goal of this research is to elucidate the effects of single mutations on the stability of the main protease (Mpro) of SARS-CoV-2.

ACM-BCB '20, Sept 21–24, 2020, Atlanta, GA

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

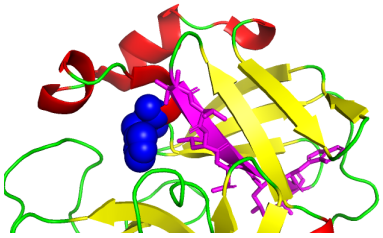

Figure 1: PDB 6y2E: Residues 25-32 (Magenta) of the main protease (Mpro) of SARS-CoV-2, in proximity to Histidine 41 (blue) which forms part of the active site.

2 MOTIVATION AND RELATED WORK

The main protease of SARS-CoV-2 is responsible for cleaving polyproteins necessary for the virus to replicate [5, 10].Residues that we focused on for mutations were those that make up a beta sheet of a canonical beta barrel close to an active site His41 at residues 25-32 [3, 4] (Figure 1). We considered this region a good candidate for mutations because such a large perturbation of a highly structured region close to the active site may prove to be highly detrimental to the protein's stability and may even affect catalytic efficiency. Various regions in the genome of the SARS-CoV-2 virus are either prone to mutations or are heavily conserved [7]. Understanding conservation and mutation rates is important, as it can offer insights about the efficacy of eventual therapeutics.

3 METHODS

Of the 127 available PDB structures of the covid-19 proteome, we selected PDB file 6y2e, of the main protease (Mpro). We assess the effects of mutating residues 25-32 exhaustively via two approaches, the online Site Directed Mutagenesis tool SDM [9], and short runs of energy minimization after generating the mutants ourselves using our in-house ProMute [1] software. SDM outputs for each mutation a predicted $\Delta\Delta G$ value, which is a measure of the extent of the effects of the mutation. Highly negative $\Delta\Delta G$ values (-5kJ/mol for example) indicate a highly destabilizing mutation, while positive values specify a highly stabilizing mutations. Near zero indicates neutral mutations. For our purposes, we didn't discern between highly stabilizing or highly destabilizing values, but only the deviation from 0 kJ/mol. We used our in-house homlogy-based ProMute software to engineer *in silico* all 152 mutants for the mentioned 8 residues, and we used the NAMD [8] Molecular Dynamics engine to energy minimize each mutant. Energy minimization was run for 500 steps, or until the mutant reached the energy level of the WT. Both the final energy and the final step needed to reach the stable energy of the WT were tallied for each mutant. We calculated standard deviation values for each metric ($\sigma_{\Delta\Delta G}$, σ_{steps} and σ_{energy}).

Res	WT	Mutant	$\Delta\Delta G$	Steps	Energy	$\sigma_{\Delta\Delta G}$	σ_{steps}	σ_{energy}
25	T	R	-0.46	301	-5492.2	0	2	0
26	T	C	0.29	497	-5428.4	1	0	1
26	T	R	-0.14	324	-5490.8	0	2	0
27	L	I	0.65	498	-5430.5	1	0	1
27	L	I	0.65	498	-5430.5	1	0	1
27	L	R	-2.86	270	-5491.5	1	2	0
28	N	A	-0.33	498	-5357.9	0	0	2
28	N	C	0.44	499	-5368.4	1	0	2
28	N	G	-1.24	499	-5363.8	0	0	2
28	N	R	-1.83	380	-5490.5	0	2	0
28	N	S	-1.65	499	-5365.7	0	0	2
28	N	V	1.81	498	-5444.8	2	0	0
28	N	W	-0.91	499	-5355.2	0	0	2
29	G	P	-3.98	498	-5431.2	1	0	1
29	G	R	-2.38	364	-5489.9	0	2	0
30	L	A	-3.84	497	-5429.3	1	0	1
30	L	D	-3.59	499	-5475.3	1	0	1
30	L	I	0.45	497	-5423.0	1	0	1
30	L	P	-4.32	481	-5491.8	2	0	0
30	L	R	-3.01	305	-5491.1	1	2	0
30	L	S	-3.71	498	-5423.8	1	0	1
31	W	R	-1.64	301	-5492.7	0	2	0
31	W	S	-3.70	499	-5432.5	1	0	1
32	L	A	-3.85	499	-5426.3	1	0	1
32	L	K	-3.34	429	-5490.6	1	1	0
32	L	R	-2.87	315	-5489.7	1	2	0
32	L	S	-3.73	499	-5429.7	1	0	1

Table 1: $\Delta\Delta G$, energy minimization step count needed to stabilize, and final energies for the most disruptive mutations. Energy=kJ/mol. A 1 or 2 in a σ column indicates a mutant is that many standard deviations from the average.

To help identify those residues which when mutated have the biggest impact on the stability of the protein, we employed a voting-type scheme to give proportional weights to metrics with higher σ values: $\text{vote}_\sigma = \sigma_1 + (1.42 \times \sigma_2)$, where σ_n refers to the amount of tests that returned results within n standard deviations of the mean. Each successive σ is weighted by the ratio between it and the threshold of the standard deviation before it (e.g. $\frac{0.98}{0.69} = 1.42$). For example, a mutation with $\sigma_{\Delta\Delta G} = 1$, $\sigma_{steps} = 0$ and $\sigma_{energy} = 1$ would have $\text{vote}_{sigma} = 2 + 0 \times 1.42 = 2.0$, while a mutation with $\sigma_{\Delta\Delta G} = 2$, $\sigma_{steps} = 1$ and $\sigma_{energy} = 0$ would have $\text{vote}_\sigma = 2.42$.

4 RESULTS AND DISCUSSION

Figure 2 lists the predicted $\Delta\Delta G$ values for the exhaustive mutations of our 8 target sites, while Table 1 lists the mutations for which at least two of the metrics were more than one standard deviation, or if one of the metrics was 2+ standard deviations from the mean.

Some residues more than others have a strong stabilizing or destabilizing effect when mutated (Figure 2). Mutating residue 31

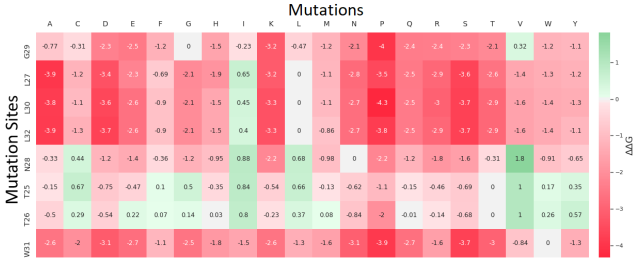

Figure 2: Heatmap of predicted $\Delta\Delta G$ value for exhaustive single amino acid mutations for residues 25-32, PDB 6y2e

always results in a $\Delta\Delta G \leq 0$, while in 75% of our loci, mutating the WT residue to Isoleucine results in a stabilizing change. Per Table 1 we see that a given mutation producing outlying data per one metric can have no bearing on the results from the other metrics. This is not surprising because different computational methods often provide complementary results. Per our vote_σ metric, the most impactful mutations were L27R, N28C, L30R, and L32R, with values of 2.4. Interestingly, 75% of these involved mutating to Arginine.

From the SDM results we observed that the most destabilizing of the sites of the main protease were W31, L32, L30, and L27. The most disruptive mutations made in these locations were to Alanine, Aspartate, Proline, and Serine. Of these combinations the one with the most disruptive prediction was a mutation to Proline at site L30. This may be due to L30 being part of a β-sheet found in domain I of the main protease, and Proline might hinder the formation of necessary hydrogen bonds with the other surrounding residues.

5 CONCLUSIONS AND NEXT STEPS

We used energy minimization relaxation curves and predicted $\Delta\Delta G$ values to assess the effects of mutating exhaustively 8 residues in the main protease of SARS-CoV-2. To determine which mutations are most impactful, we formulated a composite metric, vote_σ, and found that L27R, N28C, L30R, and L32R were the most impactful mutations. For future work we will mutate exhaustively all residues in Mpro, and not just residues 25-32, and also perform exhaustive mutations of pairs of residues within 10Å of the active site.

REFERENCES

[1] Erik Andersson and Filip Jagodzinski. 2017. ProMuteHT: A high throughput compute pipeline for generating protein mutants in silico. In *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. 655–660.

[2] Toshio Hirano and Masaaki Murakami. 2020. COVID-19: A new virus, but a familiar receptor and cytokine release syndrome. *Immunity* (2020).

[3] Zhenming Jin, Xiaoyu Du, Yechun Xu, Yongqiang Deng, Meiqin Liu, Yao Zhao, Bing Zhang, Xiaofeng Li, Leike Zhang, Chao Peng, et al. 2020. Structure of M pro from SARS-CoV-2 and discovery of its inhibitors. *Nature* (2020), 1–5.

[4] Zhenming Jin, Yao Zhao, Yuan Sun, Bing Zhang, Haofeng Wang, Yan Wu, Yan Zhu, Chen Zhu, Tianyu Hu, Xiaoyu Du, et al. 2020. Structural basis for the inhibition of SARS-CoV-2 main protease by antineoplastic drug carmofur. *Nature structural & molecular biology* 27, 6 (2020), 529–532.

[5] Hylemariam Mihiretie Mengist, Xiaojiao Fan, and Tengchuan Jin. 2020. Designing of improved drugs for COVID-19: Crystal structure of SARS-CoV-2 main protease M pro. *Signal Transduction and Targeted Therapy* 5, 1 (2020), 1–2.

[6] World Health Organization et al. 2020. Coronavirus disease 2019 (COVID-19): situation report, 134. (2020).

[7] Maria Pachetti, Bruna Marini, Francesca Benedetti, Fabiola Giudici, Elisabetta Mauro, Paola Storici, Claudio Masciovecchio, Silvia Angeletti, Massimo Ciccozzi, Robert C Gallo, et al. 2020. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *Journal of Translational Medicine* 18 (2020), 1–9.

[8] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kale, and Klaus Schulten. 2005. Scalable molecular dynamics with NAMD. *Journal of computational chemistry* 26, 16 (2005), 1781–1802.

[9] Catherine L Worth, Robert Preissner, and Tom L Blundell. 2011. SDM—a server for predicting effects of mutations on protein stability and malfunction. *Nucleic acids research* 39, suppl_2 (2011), W215–W222.

[10] Linlin Zhang, Daizong Lin, Xinyuanyuan Sun, Ute Curth, Christian Drosten, Lucie Sauerhering, Stephan Becker, Katharina Rox, and Rolf Hilgenfeld. 2020. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* 368, 6489 (2020), 409–412.

[11] Peng Zhou, Xing-Lou Yang, Xian-Guang Wang, Ben Hu, Lei Zhang, Wei Zhang, Hao-Rui Si, Yan Zhu, Bei Li, Chao-Lin Huang, et al. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *nature* 579, 7798 (2020), 270–273.