Proceedings of the 36th European Peptide Symposium
Michal Lebl (Editor)
European Peptide Society, 2022 *https://doi.org/10.17952/36EPS.2022.233*

Protocol for the Computational Optimization of Modified Peptides as Potential Protease Inhibitors

Rodrigo Ochoa and Pilar Cossio
Biophysics of Tropical Diseases, Max Planck Tandem Group, University of Antioquia, 050010 Medellin, Colombia

Introduction

From the repertoire of targets, enzymes interacting with peptide substrates constitutes one of the most interesting groups to block due to their impact as key drivers in many biological processes. Information about natural substrates as well as potential inhibitors has contributed to understand the specificity of these enzymes, and the mechanisms to recognize certain molecular binders, as in the case of proteases [1]. Many of these proteases have been identified in pathogens, or as crucial molecular components in the modulation of human diseases [2]. For example, structures of proteases have been characterized and used to design novel inhibitors of emerging viruses [3]. This is the case of the 3CL protease or main protease (Mpro) characterized for SARS-CoV-2 that since its report has motivated the publication of various studies aiming to understand its mechanism of action and screening of potential inhibitors [4,5]. Another example is cathepsin S, a protease involved in the human immunological response pathway. The overexpression of this protease is associated with some autoimmune diseases, motivating the characterization of its structure to provide new clues for designing novel binders able to control their deleterious effects [6].

Thanks to the availability of these protease structures, researchers have applied different computational strategies to design novel inhibitors, some of them mimicking the mechanism of action of their natural substrates [7]. This is the case of modified peptides composed of non-natural amino acids (NNAA), which can be associated with improved pharmaceutical properties [8,9]. These NNAAs can be computationally parameterized and included in common simulation and modelling techniques [10]. This provides valuable resources to support the engineering and design of modified peptides through computational strategies, allowing, for example, the design of novel inhibitors against proteases responsible to cleave natural peptides [11].

Something crucial is to employ computational strategies to explore efficiently both the conformational and sequence space of the peptides. In the case of the sampling techniques, there are multiple alternatives from classic to enhanced molecular dynamics simulations [12]. These trajectories can be used to score protein-peptide complexes and allow the modification of their sequences based on accepting random or guided modifications [13,14,15]. Something required is the definition of a molecular target structure bound to a peptide, which can be obtained from either a crystal structure or through docking methodologies. These targets are usually proteins that can be inhibited using peptides as potential drug-like molecules. In this work, we applied the design protocol PARCE [16] using crystal structures of two proteases of relevance, a cathepsin S and SARS-CoV-2 main protease docked with 3-mer modified peptides as starting points. The selected candidates were ranked and proposed as potential inhibitor candidates for further experimental validations.

Results and Discussion

In this work we generated a protocol to design modified peptides composed of non-natural α-L-amino acids and some basic D-amino acids. Using as input a protein-peptide complex, several mutations were attempted with a set of parameterized NNAAs. The mutations are sampled with RosettaCommons (https://www.rosettacommons.org/), and a consensus scoring or metropolis Monte Carlo criterion can be applied to accept or reject the mutations. An application using two known protease systems was included to design novel 3-mer modified peptides.

Parameterization of non-natural amino acids

Based on the information of α-L- and D-NNAAs derived from the SwissSideChain database [17], a total of 83 NNAAs were selected to perform the mutations. All the NNAAs contain chemical modifications only on the side chain, and they can be sampled using molecular dynamics (MD) simulations with the CHARMM27 force field available in SwissSideChain.

Application using proteases structures

We selected two proteases' structures from the PDB that has been co-crystallized with 3-mer modified peptides. One is the SARS-CoV-2 Mpro structure (PDB id 6fv2) [18], and the second is a cathepsin S structure (PDB id 2f1g) [19]. In both cases, the bound peptides are composed of NNAAs that are not part of the parameterized NNAA dataset. Based on that, we modelled a group of 3-mer peptides using combinations from a set of 14 NNAAs available in RDKit (https://rdkit.org/) to run conformer predictions. The conformer library was docked using AutoDock Vina [20], and the bound peptides with the best energies were used as starting points for the design of modifications. In the case of Mpro, the sequence was DGN (D-glutamine), ORN (ornithine) and DME (D-methionine). For cathepsin S, the 3-mer was DAR (D-arginine), ORN (ornithine) and DPR (D-proline). A graphical representation of the structures and the docked poses are shown in Figure 1.

Both complexes were used as references to design variants with higher affinities. For the design, we initially attempted a run using a single scoring function with a Metropolis Monte Carlo criterion to accept the mutations. A total of 400 mutations were attempted, and the score evolution was monitored to identify the accepted candidate sequences. We found a high number of accepted sequences (85)

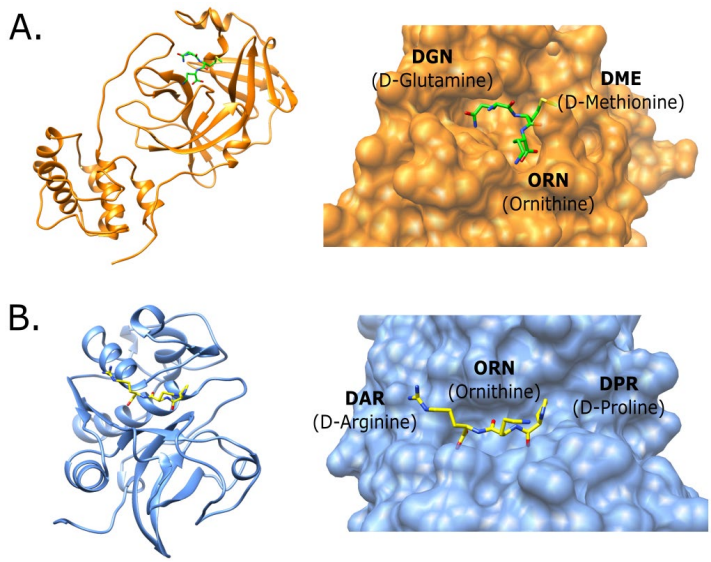

Fig. 1. Structures of the two proteases that were included for the application. (A) Structure of a SARS-CoV-2 Mpro structure (PDB id 6fv2) bound to the starting 3-mer peptide DGN (D-glutamine), ORN (ornithine) and DME (D-methionine. (B) Structure of the cathepsin S structure (PDB id 2f1g) bound to the 3-mer peptide DAR (D-arginine), ORN (ornithine) and DPR (D-proline).

from the 400 iterations (see Figure 2). However, relying in one single scoring function can be risky given the intrinsically errors associated [21]. Because of that, the next step was to run the design with a consensus setup, using two thresholds of three and four scoring functions to accept the mutations from the original six.

After running the design with a threshold of three and 200 mutation attempts, we found more difficulties to converge the scoring functions after the planned number of iterations. This is mostly given by the small size of the peptide sequences, which can change drastically the average scores from one iteration to the next. For example, we obtained for the Mpro system a total of 20 candidates, but with a noticeable lack of convergence. For the cathepsin S, the results were similar with a total of 34 accepted molecules. To avoid these issues, we increased the threshold from three to four scoring functions to accept the mutations. In that scenario, a lower number of sequences were accepted, but we got a better performance of the scoring functions during the iterations. An example with four of the

six scoring functions values for the accepted sequences during the 200 mutation attempts is shown in Figure 3.

We found that from the six scoring functions, only NNscore [22] finished the design run with sequences having more positive scores with respect to the original molecules. This is expected given the stochastic nature of the sequence space search, where the scores can also increase in energy to overcome local barriers. Also, by setting a higher threshold, the number of accepted sequences was lower in both cases, with a total of six peptides for each system. Despite the numbers, there are higher probabilities to obtain candidates with better average scores than the reference.

The NNAAs that are part of the prioritized molecules are: DGN (D-glutamine), DME (D-methionine), ALO (allo-threonine), BB8 (phenylserine), C2N (3-chloro-l-alanine), PF5 (pentafluoro-phenylalanine), HOX (4-amino-L-phenylalanine), MEG ((3s)-3-methyl-l-glutamic acid), DLY (D-lysine), DAL (D-alanine), ALC (3-cyclohexyl-alanine), WFP (3,5-Difluoro-phenylalanine), 2FM (s-(difluoromethyl)-homocysteine), ABA (2-Aminobutyric acid), PHI (4-iodo-phenylalanine) and TYI (3,5-diiodotyrosine).

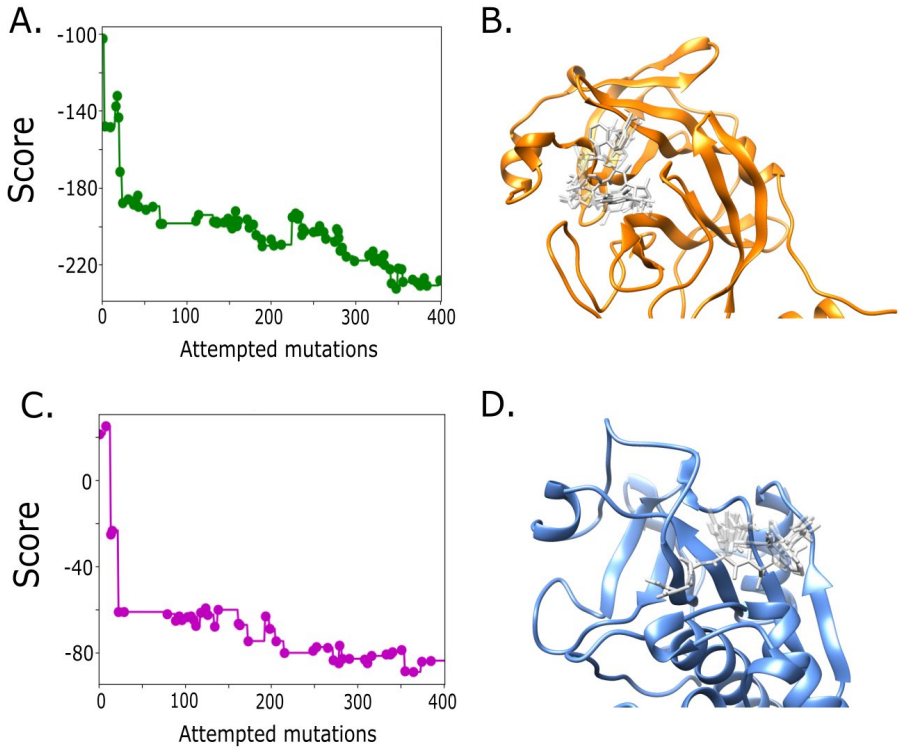

Fig. 2. Evolution of the Rosetta scoring function after 400 mutation attempts using a Metropolis criterion for Mpro (A) and cathepsin S (C). The dots in the curve represent the mutations that were accepted. A structural representation of the accepted modified peptides is shown for Mpro (B) and cathepsin S (D).

Finally, to show the connection between our design protocol and the possibility to implement MD simulations to re-rank the candidates, all the peptides were subjected to 100 ns MD with the modified CHARMM27 force field available at the SwissSideChain database. The trajectories were scored using the same six scoring functions of the design, and an average ranking was calculated to prioritize the candidates. As a perspective, the user has the possibility to re-rank the molecules using more exhaustive sampling methodologies before moving to experimental validations.

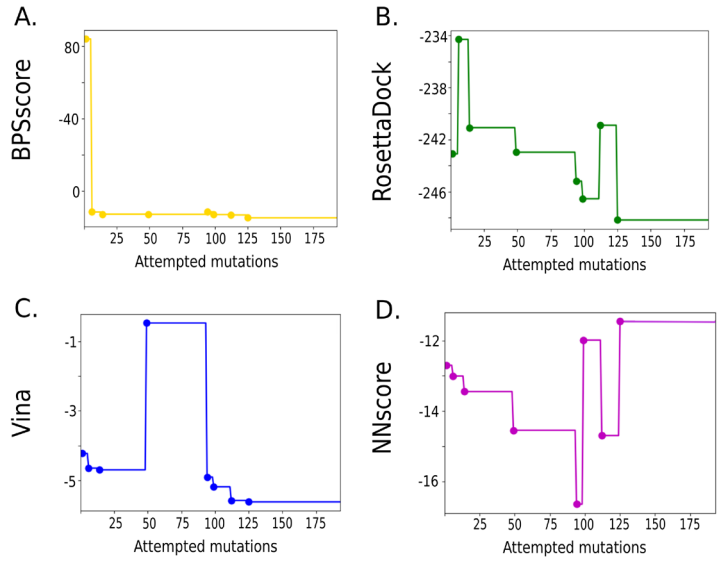

Fig. 3. Evolution of some of the scoring functions using a consensus criterion. We used a total of six scoring functions to calculate the consensus with a threshold of four after attempting 200 mutations. The dots in the curve represent the mutations that were accepted. The shown scoring functions are (A) BPSScore, (B) RosettaDock, (C) Vina and (D) NNscore.

Acknowledgments

This work, R.O. and P.C. have been supported by MinCiencias, University of Antioquia and Ruta N, Colombia, the Max Planck Society, Germany.

References

1. Ochoa, R., et al. *BMC Bioinformatics* **21**, 1-20 (2020), https://doi.org/10.1186/s12859-020-03931-6
2. Drag, M. and Salvesen, G.S. *Nat. Rev. Drug Discovery* **9**, 690-701 (2010), https://doi.org/10.1038/nrd3053
3. Zhou, Y., et al. *Antiviral Res.* **116**, 76-84 (2015), https://doi.org/10.1016/j.antiviral.2015.01.011
4. Zhang, L., Lin, D., Sun, X., et al. *Science* **368**, 409-412 (2020), https://doi.org/10.1126/science.abb3405
5. Macchiagodena, M., et al. *Chem. Phys. Lett.* **750**, 137489 (2020), https://doi.org/10.1016/j.cplett.2020.137489
6. Tato, M., Kumar, S., Liu, Y., et al. *Sci. Rep.* **7**, 1-15 (2017), https://doi.org/10.1038/s41598-017-01894-y
7. Mulligan V.K., et al. *Proc. Nat. Acad. Sci.* **118**, e2012800118 (2021), https://doi.org/10.1073/pnas.2012800118
8. Gentilucci, L., et al. *Cur. Pharm. Des.* **16**, 3185-3203 (2010), https://doi.org/10.2174/138161210793292555
9. Vagner, J., et al. *Cur. Opin. Chem. Biol.* **12**, 292-296 (2008), https://doi.org/10.1016/j.cbpa.2008.03.009
10. Lenci, E. and Trabocchi, A. *Chem. Soc. Rev.* **49**, 3262-3277 (2020), https://doi.org/10.1039/D0CS00102C
11. Fear, G., et al. *Pharmacol. Therap.* **113**, 354-368 (2007), https://doi.org/10.1016/j.pharmthera.2006.09.001
12. Miao, Y. and McCammon J.A. *Mol. Simul.* **42**, 1046-1055 (2016), https://doi.org/10.1080/08927022.2015.1121541
13. Ochoa, R., et al. *J. Chem. Inf. Mod.* **59**, 3464-3473 (2019), https://doi.org/10.1021/acs.jcim.9b00403
14. Ochoa, R., et al. *Front. Mol. Biosci.* **8**, 636562 (2021), https://doi.org/10.3389/fmolb.2021.636562
15. Ochoa, R., et al. *Frontiers in Immunology* **13**, 862851 (2022), https://doi.org/10.3389/fimmu.2022.862851
16. Ochoa, R., et al. *Comp. Phys. Commun.* **260**, 107716 (2021), https://doi.org/10.1016/j.cpc.2020.107716
17. Gfeller, D., et al. *Nucl. Acids Res.* **41**, D327-D332 (2012), https://doi.org/10.1093/nar/gks991
18. Zhang, L., et al. *J. Med. Chem.* **63**, 4562-4578 (2020), https://doi.org/10.1021/acs.jmedchem.9b01828
19. Tully, D., et al. *Bioorg. Med. Chem. Lett.* **16**, 1975-1980 (2006), https://doi.org/10.1016/j.bmcl.2005.12.095
20. Trott, O. and Olson, A.J. *J. Comput. Chem.* **31**, 455-461 (2010), https://doi.org/10.1002/jcc.21334
21. Michel, J. and Essex, J.W. *J. Comp. Aided Mol. Des.* **24**, 639-658 (2010), https://doi.org/10.1007/s10822-010-9363-3
22. Durrant, J.D. and McCammon, J.A. *J. Chem. Inf. Mod.* **51**, 2897-2903 (2011), https://doi.org/10.1021/ci2003889