Virtual screening of peptides with high affinity for SARS-CoV-2 main protease

William Farias Porto [1]

Porto Reports, Brasília, DF, 72236-011, Brazil

ARTICLE INFO

Keywords:
Genetic algorithm
RESTful API
Molecular docking
COVID-19

ABSTRACT

The current pandemic of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused more than 2,000,000 deaths worldwide. Currently, vaccine development and drug repurposing have been the main strategies to find a COVID-19 treatment. However, the development of new drugs could be the solution if the main strategies fail. Here, a virtual screening of pentapeptides was applied in order to identify peptides with high affinity to SARS-CoV-2 main protease (M^{pro}). Over 70,000 peptides were screened employing a genetic algorithm that uses a docking score as the fitness function. The algorithm was coupled with a RESTful API to persist data and avoid redundancy. The docking exhaustiveness was adapted to the number of peptides in each virtual screening step, where the higher the number of peptides, the lower the docking exhaustiveness. Two potential peptides were selected (HHYWH and HYWWT), which have higher affinity to M^{pro} than to human proteases. Albeit preliminary, the data presented here provide some basis for the rational design of peptide-based drugs to treat COVID-19.

1. Introduction

Since December 2019, the worldwide massive health crisis provoked by the newly identified severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused over 2,000,000 deaths worldwide, according to World Health Organization.[2] As a global imperative, vaccine development has been accelerated, with more than 100 vaccine candidates [1]. Although there are three FDA-approved vaccines,[3] from Moderna, Pfizer-BioNTech and Janssen, SARS-CoV-2 is still a serious issue to combat and investigate at the molecular level [4]. Some aspects of coronavirus disease 2019 (COVID-19) remain unknown, mainly due to the emergence of novel strains [5].

The SARS-CoV-2 RNA genome consists of about 30,000 bp. The largest open read frame, ORF1a/b, encodes polyproteins 1a and 1 ab (pp1a and pp1ab) [6–9]. These proteins are further processed by the main protease (M^{pro}) and a papain-like protease to produce different functional proteins [6,10]. This cleavage releases important viral enzymes, including RNA-dependent RNA polymerase, helicase and methyltransferase [6]. Therefore, M^{pro} is key to the viral cycle and, for this reason, this enzyme has been proposed as a therapeutic target for anti-coronavirus drug development [11,12,36,37].

Drug repurposing has been extensively explored as an attempt to identify possible M^{pro} inhibitors, where docking experiments play a critical role in virtual screening projects [13–15]. Drug repurposing is an elegant approach, because these drugs are already approved for human use [16]. However, the only officially approved drug for COVID-19 treatment is remdesivir for very specific cases [17]. Therefore, considering a pessimistic scenario, anti-SARS-CoV-2 drugs need to be developed from scratch.

Peptide therapeutics have gained attention in the last decade [18–20], with some recent possible applications for COVID-19 [21,22]. These peptides offer a wide combinatorial space to explore (*i.e.* for a decapeptide there are 20^{10} possible combinations, taking into account only the 20 proteinogenic amino acid residues), and this enormous combinatorial space allows the development of inhibitors for different enzymes [20,23–25]. Besides, there is no convergence between different techniques for yielding such peptides, allowing different solutions for the same problem [20]. The main *in vitro* technique has been the high throughput screening of chemical, genetic and/or recombinant libraries, which could explore about 10^8-10^{13} different peptides [19]. For the *in*

silico counterpart, virtual screening is the alternative mean to identify possible peptide therapeutics, using docking as the main engine [26–28]. Therefore, M^{pro} inhibitors based on peptides could be an alternative for COVID-19 treatment.

In fact, computer science and technology information applications have contributed in different ways to dealing with the pandemic [29]. Drug repurposing has been the main application of virtual screening; however, this technology could also be applied for exploring the combinatorial peptide space. Therefore, here, a virtual screening

strategy using docking and genetic algorithms, speeded up by information technology applications, was developed to identify peptides with high affinity to M^{pro}. Two peptides with high affinity to M^{pro} were identified, and their possible applications to develop new drugs to treat COVID-19 are discussed.
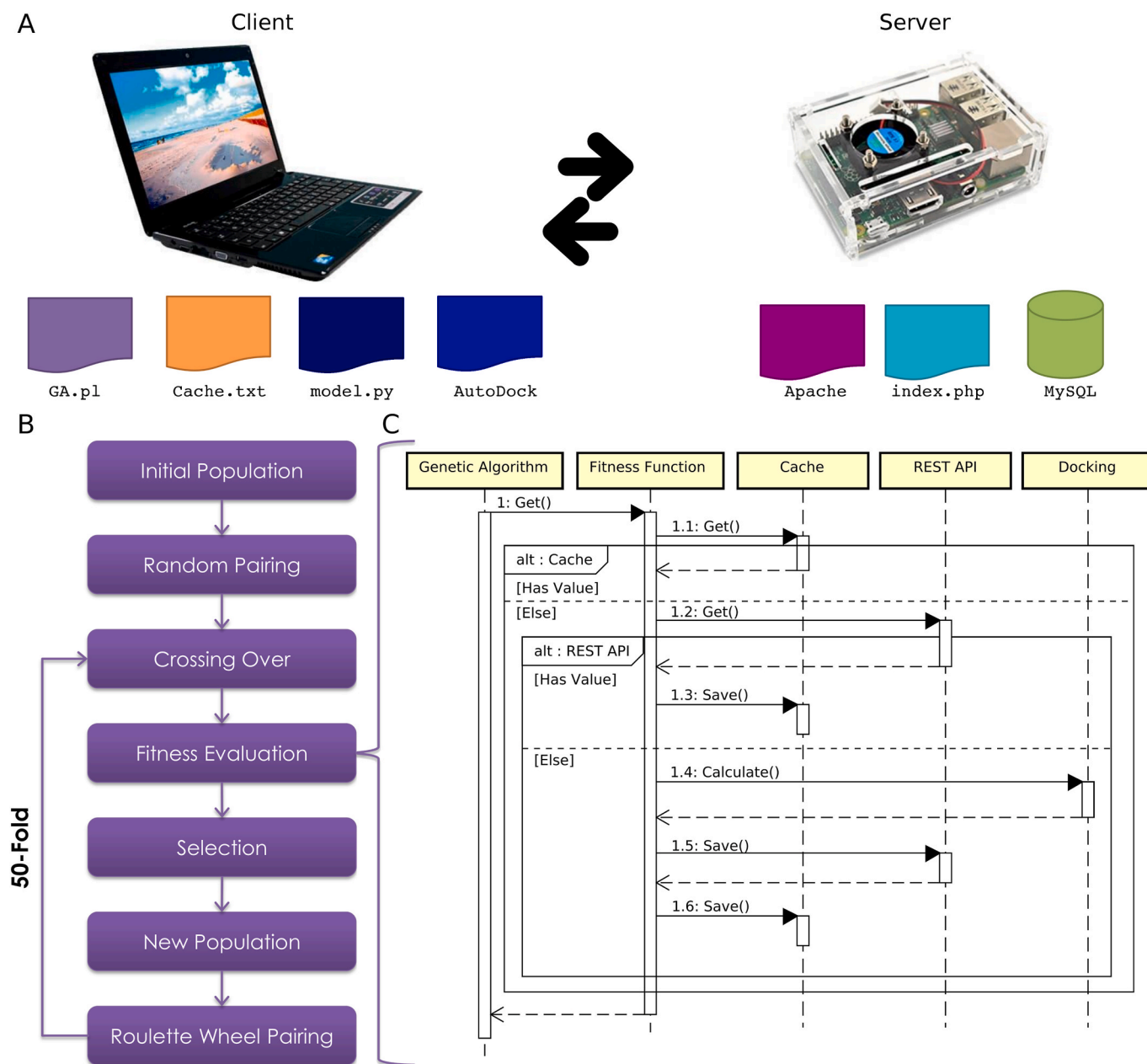

Fig. 1. Virtual screening system architecture scheme. (A) Client-server architecture. On the client side, the application was composed of a genetic algorithm, in PERL, a cache file, a 3D modelling script in python, using PyMOL interface and AutoDock Tools and AutoDock Vina, both simplified as AutoDock. Eight Intel i7 cores plus three raspberry pi 3 cores were used as independent clients; other client instances were used occasionally. On the server side, a raspberry pi 3 was used, running the LAMP stack due to its lower computing power compared to Intel i7 processor. A RESTful API was developed to persist the peptide data, reducing the time of processing docking experiments. (B) Genetic algorithm flowchart. The 19 pentapeptides were used as the initial population; in the first iteration a totally random sequence pairing system for crossing over was applied, in order to improve the diversity of sequences and in the subsequent iterations a roulette wheel pairing model was applied for selection of sequences for crossing over. (C) Fitness function sequence diagram. This function was developed to reduce the need for docking processing. Firstly, the algorithm tries to get the information in cache file; if the data exists, it is returned to the genetic algorithm; otherwise, the RESTful API is triggered; if the data exists, it is returned to fitness function, saved in cache, and returned to the genetic algorithm; otherwise, docking process is required to create the data, which is returned to the fitness function, saved in RESTful API and in cache and finally returned to the genetic algorithm.

2. Results

2.1. Minimum exhaustiveness dockings for a huge number of peptides

The virtual screening system was constructed using a client-server architecture, which allowed the task distribution in different computers and/or different cores of multicore processors and, due to the persistence layer on server side, more than 70,000 peptide sequences were explored (Fig. 1). The structure of SARS-CoV-2 M^{pro} was used as the target for developing peptides with high affinity by means of the genetic algorithm. The genetic algorithm simulates the evolution of a set of sequences, the population, by a number of generations. Thus, the population of peptide sequences was evolved using the docking scores against the M^{pro} active site, increasing the score and, therefore, increasing the affinity of these peptides to the enzyme.

Fig. 2 shows the overall assessment of our virtual screening system. Due to the prevalence of aromatic residues, the same simulation excluding those residues was performed; however, none of them reached the affinity of the complete set (Fig. 2A). In fact, the rarefaction curves (Fig. 2B) indicated that there are more sequences to be discovered as more independent simulations are performed for both amino acid sets; however, the aliphatic-only set is more diverse than the full set. This effect should occur due to the preference for aromatic amino acids (including histidine), as demonstrated by Fig. 2A. Therefore, there was a small overlap between the sets, indicating that the aromatic amino acids rapidly populate the full set simulations (Fig. 2C).

Although the algorithm generated evolved populations, final sequences from each population were not selected, instead, the database was used to select those sequences with high affinity to M^{pro}. Sequences with scores higher than 8955 (-9.1 kcal mol^{-1}) were subjected to cluster analysis, in order to select the best sequence from each cluster for further analysis. Three clusters were selected (Fig. 3). Overall, sequences presented at least one Trp and one His residue. In addition, the best sequence from the simulation without aromatic side chains was also selected. Thus, the number of compounds was reduced from more than seventy thousand to only four peptides.

2.2. Medium exhaustiveness dockings for a reduced number of peptides

With fewer compounds to analyze, more exhaustiveness docking could be applied. Due to the fact that the initial dockings were performed with minimum exhaustiveness, the results were coarse and then
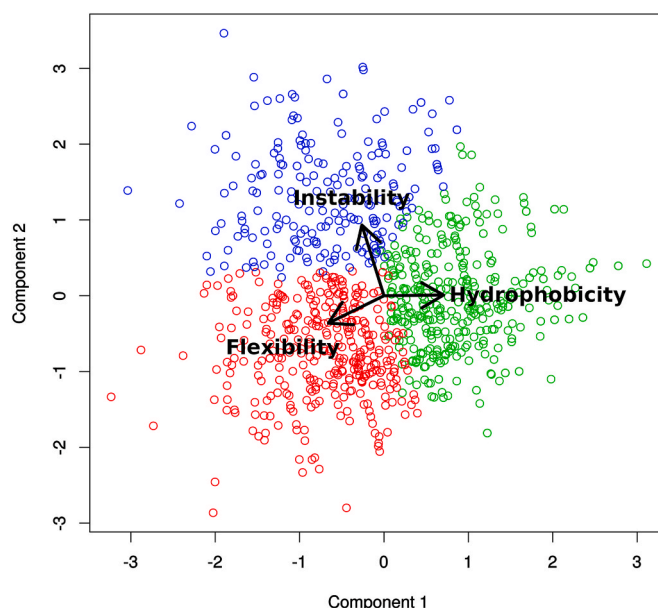

Fig. 3. Principal component analysis and cluster diagram from sequences generated by genetic algorithm. Only the sequences with high fitness values were selected. Three physicochemical properties (hydrophobicity, flexibility and instability) were used to define the three peptide clusters, where the sequence with highest fitness value was selected as the prototype for next analysis.

they were refined. Besides, some residue combinations might not be tested by the genetic algorithm (in particular, proline residues were not included). Thus, amino acid scanning was performed. The scanning indicated that some point modifications improved the docking score (Fig. 4).

Although point modifications in different positions increased the docking score, combining two or more of them resulted in reduced scores. Both combination models, simply-the-best and Joker, did not achieve better scores than the original sequences (data not shown). Therefore, from amino acid scanning, the sequence with the highest affinity or a value close to it (polar amino acids were preferred due to water solubility) was selected for further analysis, resulting in the sequences WWHWR, HHYWH, HYWWT and ARAHR.
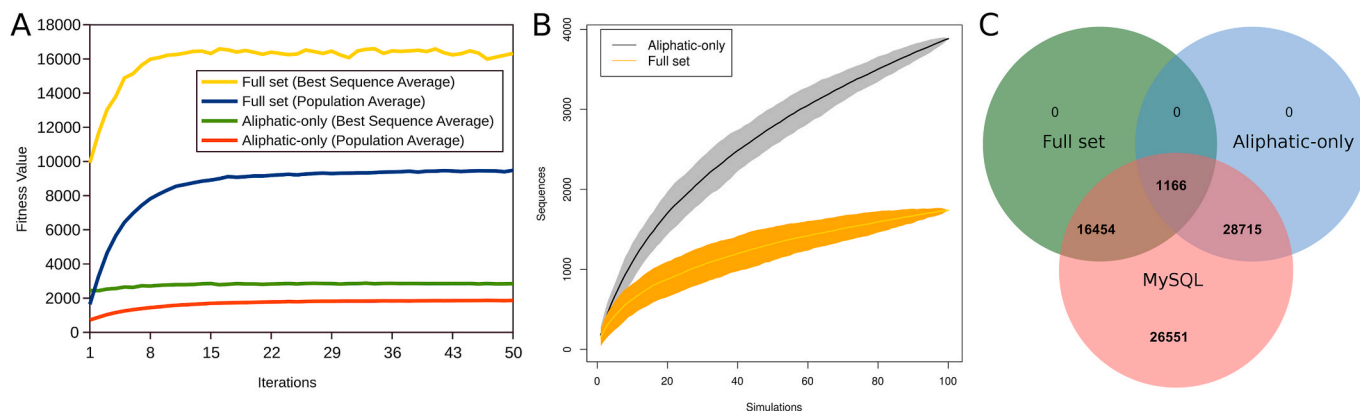

Fig. 2. Genetic Algorithm outcome. (A) Fitness evolution along the algorithm iterations. The best sequence and population averages for both sets are displayed. The full set, with aromatic amino acid residues, reached higher fitness values than the aliphatic-only set. While the fitness function in full set increased both in population and best sequence, in the aliphatic-only, there is only a smooth increase, which could be related to the fact that M^{pro} is chymotrypsin-like and the aromatic residues are needed to increase the peptide-enzyme affinity. (B) Peptide rarefaction curves. Sequences were retrieved from iterations 25–50 in all 100 independent simulations. In both cases, the curve indicates more peptides could be found if more simulations were performed. However, the aliphatic-only set is more diverse than the full set. (C) Venn diagram of generated sequences. The full and aliphatic-only sets represent 100 complete independent simulations, while MySQL indicates the sequences that persisted in the database. The aliphatic-only set has almost twice as many sequences as the full set, and the remaining sequences present only in MySQL are results of incomplete simulations or prototype experiments.
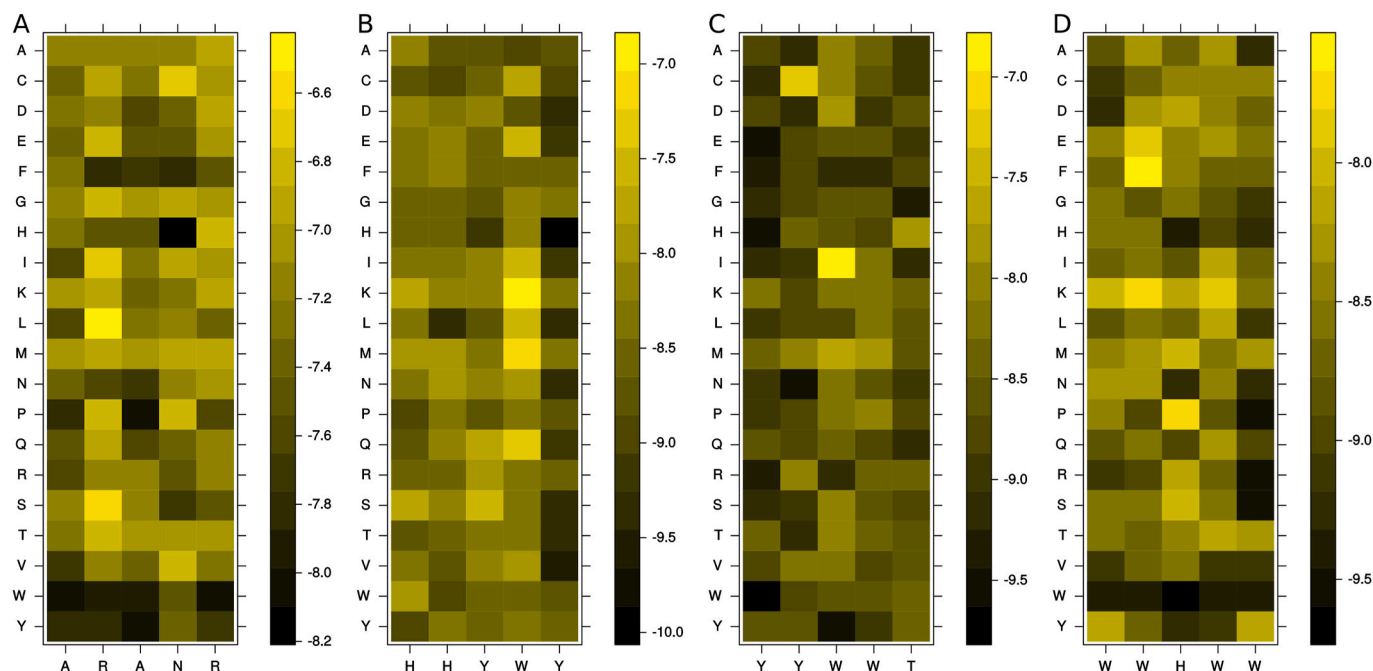
Fig. 4. Amino acid scanning of prototype peptides. Each position was replaced by one of twenty proteinogenic amino acids, generating a variant, which was subjected to high exhaustiveness dockings in triplicate, and the results are shown in average values. (A) The aliphatic-only sequence; (B, C and D) the prototypes of instability, flexibility and hydrophobicity clusters, respectively. Overall, the inclusion of aromatic residues tends to improve the docking scores, while the modifications in the three-central residues are less tolerated.

2.3. High exhaustiveness dockings for a limited number of peptides

Given that the designed molecules are peptides and the target is a protease, we need to verify whether the peptides could bind to human proteases with more affinity than the virus protease. Thus, designed peptides were subjected to maximum exhaustiveness dockings with 100-fold repetitions. The peptide WWHWR showed no statistical differences in its scores against M^{pro} and human chymotrypsin, and thus this peptide would not be suitable for drug development. The remaining peptides showed higher affinity to the viral protease than to human proteases (Fig. 5).

Binding poses from dockings on M^{pro} indicated that the peptides assume a similar conformation to the 11a inhibitor. The indole ring from the 11a inhibitor is overlapped by Trp^4 on HHYWH and HYWWT

(Fig. 6A), indicating that both sequences present the "HYW" motif; however, the 4th residue is key to the correct positioning in M^{pro} active site. The M^{pro} residues Thr^{25}, Thr^{26}, His^{41}, Cys^{44}, Glu^{166}, Pro^{168} and Arg^{188} showed interactions with both peptides (Fig. 6B and C). The 11a inhibitor is smaller than the pentapeptides, where there is no overlap between the first two residues and the 11a inhibitor (Fig. 6A).

3. Discussion

The COVID-19 pandemic has provided the opportunity to apply the basic science developed in the last few decades and also to develop some frugal innovations [30]. Numerous efforts to reposition drugs against SARS-CoV-2 have been made. Indeed, virtual screening has been the leading strategy for selecting drug candidates for repurposing against
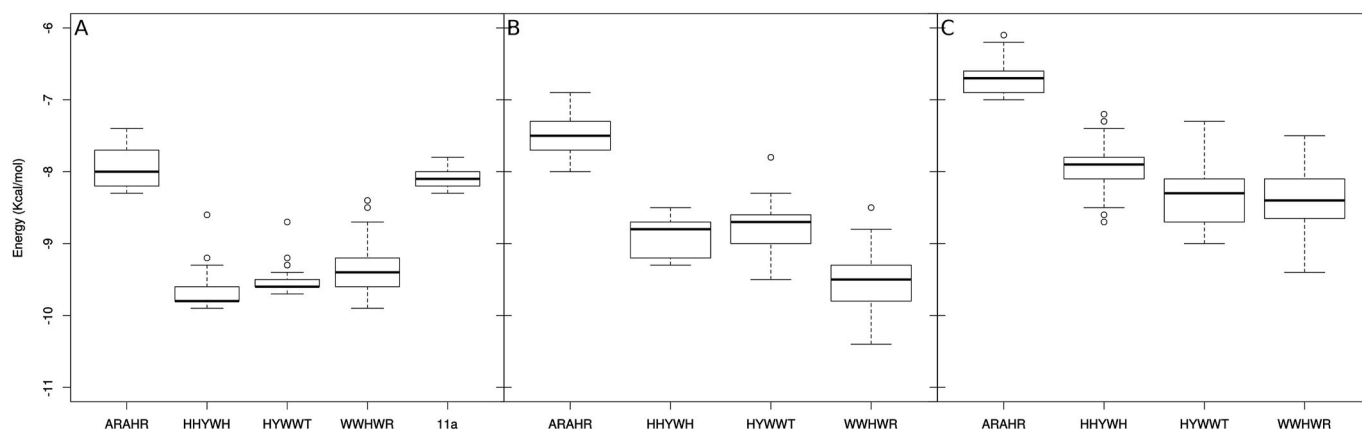
Fig. 5. Energy from maximum exhaustiveness docking with 100-fold repetitions. The four peptides were docked against (A) SARS-CoV-2 M^{pro}, (B) human chymotrypsin and (C) human trypsin. Each docking was repeated 100 times using different seeds. For M^{pro}, an additional docking experiment was performed with the inhibitor 11a, from the original structure. Overall, the peptides demonstrated lower affinity to human trypsin than human or viral chymotrypsin. The peptide WWHWR showed higher affinity to human chymotrypsin than M^{pro} (p-value = 0.00047), while the peptides HHYWH and HYWWT showed higher affinity to M^{pro} than human chymotrypsin (p-values<2.2e-16 for both). The ARAHR peptide showed less affinity to M^{pro} than 11a inhibitor (p-value = 3.317e-05).

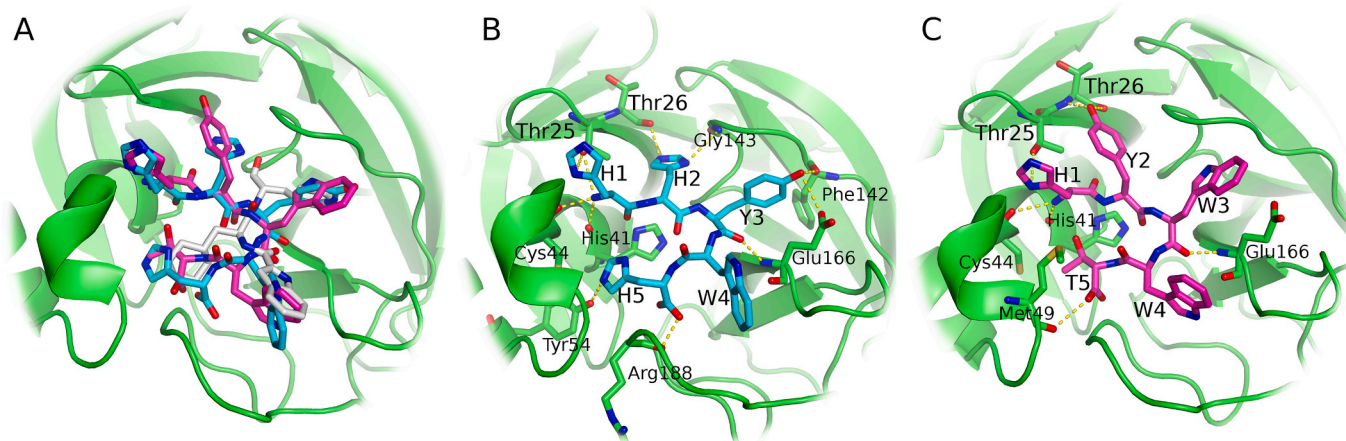**Fig. 6.** Docking poses from HHYWH and HYWWT peptides. (A) The overlap of both peptides with 11a inhibitor. (B and C) Electrostatic interactions between peptides (B) HHYWH and (C) HYWWT and M^{pro}. The interactions are indicated by yellow dotted lines. Residues from M^{pro} are labeled with three-letter code; residues from pentapeptide are labeled with single-letter code. The majority of interactions involve the main M^{pro} main chain, instead of side chains, except for Thr^{25}, Tyr^{54} and Glu^{166} and HHYWH; and Thr^{25} and HYWWT. The list of all interactions is available in Supplementary Table 1.

COVID-19 [13,14].

Drug repurposing is a fast and elegant approach, because the drug is already approved for human use. However, the current scenario indicates that this could be a dead-end and new drugs should be developed. Indeed, a vaccine may be the ultimate resource to prevent SARS-CoV-2 infection, but the COVID-19 mechanisms are not well enough understood [31,32]. The emergence of other SARS-CoV-2 strains [5] makes it necessary to carry out continuous screening to find new possible solutions. Besides, protective immunity may only last for short time periods [33].

In the viral cycle, viral proteases play a critical role in processing the viral polyprotein, and they were therefore used as targets for antiviral therapy, with a classical example of HIV treatment based on protease inhibitors [34].

The development of peptide inhibitors targeting key enzymes in a pathogen's metabolism has gained attention during the last decade. This situation is clearly depicted when it comes to β-lactamase enzymes, which confers resistance to bacteria, with several peptide inhibitors being developed by using different strategies [20].

Despite the enormous combinatorial space of pentapeptides (19^5, excluding proline residues), the strategy here applied (Fig. 1) was able to explore the combinatorial space with an optimal performance (Fig. 2). Although combinatorial space for pentapeptides was 19^5 combinations, a target number was not defined in the algorithm; instead, this number was used as an open goal, and when this goal was reached, the goal could be doubled. Indeed, Fig. 2B indicated that more simulations are needed to reach a plateau in the discovery of pentapeptides.

In the first virtual screening step, more than 70,000 pentapeptide sequences were screened (Fig. 2C), which would require high computational power. In this context, the minimum exhaustiveness dockings and the information technology application speeded up the virtual screening process. In fact, information technology played a crucial role in this work, because the system architecture using a server application allowed fast screening using the genetic algorithm, given that an HTTP request is faster than a docking experiment. Also, despite Moore's law reaching its limit [35], the client-server architecture of the whole application allowed the raspberry pi to simulate about 10% of all genetic algorithm simulations, even with less processing power. Nevertheless, the minimum exhaustiveness dockings had a cost: the noise in data could result in a non-optimal choice of molecules. To overcome this bias, the next steps were designed to exclude the noise, using the cluster analysis (Fig. 3), amino acid scanning with maximum exhaustiveness (Fig. 4) and several replicates. At this point, the use of docking replicates is important because peptides are flexible ligands, and docking simulations could

therefore explore this property. The resulting peptides showed affinities in the range between −10 and −9 kcal mol^{-1} (Fig. 5), which is in a similar range as that of lichen metabolites [36]; and a series of compounds from ChEMBL [37] against M^{pro}. Although some compounds from other works presented affinities below −10 kcal mol^{-1} [36,37], a direct comparison should be made with care, due to the differences in the number of docking replicates. Nevertheless, in comparison with the 11a inhibitor, under the same conditions, this approach resulted in peptides with higher affinities to M^{pro} (Fig. 5). Furthermore, the peptides should have some degree of selectivity towards M^{pro}, with less probability of binding to human proteases (Fig. 5). Therefore, the prototype of hydrophobicity cluster was discarded, due to its predicted affinity to human chymotrypsin.

However, upon binding, these peptides could inhibit the enzyme or they could be cleaved by the enzyme. In fact, in both situations, they can be useful. In the case of inhibition, the peptide could be linked to a cell-penetrating peptide (CPP) [22] and inhibit the virus's replication; otherwise, a toxin could be designed to kill the infected cells, by a combination of a four-domain peptide, including a toxin, the peptide, a toxin inactivating sequence and a CPP. Once cleaved on the pentapeptide, the toxin would be released and the cell would die. Due to the selectivity towards viral protease, only the infected cells would be affected. Another possibility is the use of these peptides as scaffolds for peptidomimetic [38] or nucleopeptide [39] development, which would take advantage of their high affinities to M^{pro} and inhibit viral replication.

4. Conclusions

Peptides have been considered "the drugs of the future" and, although they are preliminary, the data presented here provide some basis for the rational design of peptide-based drugs to treat COVID-19. The actual activity of such peptides remains unknown. However, independently of whether they act by inhibition or just by binding with high affinity to SARS-CoV-2 M^{pro}, these peptides may represent only part of the puzzle solution. As discussed, they could be used for engineering multi-domain peptides with different mechanisms depending on the core activity (inhibition or just binding) or even for application as part of a drug cocktail, similar to HIV treatment. Despite the development of vaccines against COVID-19 [40], these peptides could be useful as a treatment for those who are hospitalized with severe COVID-19 infection.

5. Material and methods

5.1. Virtual screening system architecture

A raspberry pi 3 running the LAMP stack (Linux, Apache, MySQL and PHP) was used as the server layer. This was used for persisting data about peptides generated by genetic algorithm. A RESTful API was developed, allowing the communication between the server and the clients, where the execution of the genetic algorithm was performed. The GET and POST HTTP protocols were used to retrieve or post the data, respectively. This layer allowed maximum parallelization and data integration between different computers acting as clients.

On the client layer, the local cache file is the primary resource of fitness data; when the value was not found in cache, the RESTful API was triggered, where for determining the fitness function, the algorithm sent a GET request to verify the existence of this registry; if the registry existed, the JSON was processed to get the fitness value; otherwise, the fitness had been calculated by the client and then sent using POST request for data persistence. After retrieving the data either by docking or RESTful API, the value was added to the local cache file. Eight Intel i7 cores plus three raspberry pi 3 cores were used as clients, running independent simulations.

5.2. Genetic algorithm

The genetic algorithm simulates the evolution of a population of sequences during a number of iterations, where given iteration I_n generates the population P_n from the population P_{n-1}, evaluating the sequences according to the value of a fitness function, also known as chance of survivor and mating (Fig. 1B). Here the algorithm was implemented in PERL programming language, according to Porto et al. [50]. In the first iteration (I_1) of the implementation of our custom genetic algorithm, P_0 was composed by 19 pentapeptides, each one for a respective amino acid residue (e.g. AAAAA, RRRRR or WWWWW); proline residues were not included due to a bug in automatic structure generation. All sequences from P_0 had the same fitness value, thus providing a random pairing for crossing over (Fig. 1). From iteration I_2 to I_n, the sequence pairing for mating was performed according to the corresponding fitness values by means of a roulette wheel pairing model. For each iteration, 50 sequence pairs were selected from population P_n and each pair was submitted to a crossing over process, generating a couple of children for population P_{n+1}. No mutation was applied. Next, sequences from P_{n+1} were evaluated by fitness function. The 10 worst sequences were removed from the population P_{n+1} and then another iteration step was initiated (Fig. 1B). The cycle was repeated until the number of iterations was exhausted. One hundred independent simulations were performed, each one with 50 iterations using the same conditions. A second set of simulations was performed under the same conditions, excluding His, Phe, Tyr and Trp residues, because these residues seem to increase the affinities between peptide and enzyme.

5.3. Fitness function

The fitness value was obtained either from the persisted data (cache or RESTful API) or by calculating the fitness value as a function of a docking score, by means of Eq. (1):

$$fitness = e^{|DS|} \tag{1}$$

where the *DS* states the docking score, which was given by AutoDock Vina (see below). The modulus of docking score was used in order to increase the fitness value, while the exponential function was applied because the docking score has only one decimal place, which could not be reflected in great fitness differences in the roulette wheel selection from the genetic algorithm. For the automatic docking subroutine, an extended peptide structure had been generated by a PyMOL script,

before performing the minimum exhaustiveness docking using Auto-Dock Tools [41] and AutoDock Vina [42].

5.4. Docking

AutoDock Vina [41] and AutoDock Tools [42] were used to perform molecular docking, which was used to verify the affinities between the generated peptides and SARS-CoV-2 M^pro (PDB ID: 6LZE) [43] or human proteases (PDB IDs: 4H4F and 1TRN) [44,45]. The PDB structures were converted to PDBQT files using AutoDock Tools. Dockings were performed using AutoDock Vina. For all structures, the box size was set as 90 Å^3; then the grid box was positioned in X, Y, Z coordinates according to the respective enzyme: 24.33, 22.34 and −4.27 for 1TRN; 28.86, 44.46 and 16.10 for 4H4F and −10.94, 12.69 and 68.91 for 6LZE. Depending on the virtual screening step, minimum or maximum exhaustiveness were used.

5.4.1. Minimum exhaustiveness dockings

Minimum exhaustiveness dockings were performed by setting the AutoDock Vina exhaustiveness and the number of CPUs to use options as one. This allowed maximum parallelization of independent docking experiments. No replicas were performed for these dockings.

5.4.2. Maximum exhaustiveness dockings

Maximum exhaustiveness dockings were performed using the AutoDock Vina default options. Dockings were performed with three or one hundred independent runs with different seeds, according to the virtual screening step.

5.5. Clustering

The sequences with fitness values higher than 8955 (corresponding to docking scores below −9.1), were selected for cluster analysis according to physicochemical properties. Three physicochemical properties were calculated: average hydrophobicity using the Wimley & White scale [46]; average flexibility using the Bhaskaran-Ponnuswamy scale [47]; and peptide instability using the Guruprasad scale based on dipeptide compositions [48]. Three clusters were defined by PAMK algorithm, using the R package for statistical computing. The best sequence from each cluster was selected for the next steps. The best sequence from the simulations without the aromatic ring containing residues was also selected for the next steps.

5.6. Amino acid scanning

The four sequences selected from virtual screening were subjected to amino acid scanning, where each position was replaced by each proteinogenic amino acid residue, generating a variant with a single amino acid alteration. MODELLER 9.23 [51] was used to build a model from each primary sequence, by means of environ and model classes. These variants were subjected to maximum exhaustiveness dockings in triplicate.

5.7. Combinations

Two models for combining data from amino acid scanning were assumed. In the first model (simply-the-best), each position was filled up by the amino acid residue with the highest affinity; while in the second model (Joker), a prosite grammar was constructed using the amino acid residues with higher affinity than the native one, and then the Joker algorithm [49] was used to select the residue according to the hydrophobicity of each position in the prosite grammar. MODELLER 9.23 [51] was used to build a model from each primary sequence, by means of environ and model classes. Combinations were forwarded to maximum exhaustiveness docking experiments in triplicate. The Joker combinations were used as a control because their hydrophobic scale favors the

selection of aliphatic instead of aromatic hydrophobic residues.

5.8. Statistical analysis

The one-sided Mann-Whitney-Wilcoxon test was used to verify the differences between the affinities among the peptides and 11a inhibitor towards M^{pro}; and among the M^{pro} and human proteases for each peptide. The test was performed using the R package for statistical computing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi. org/10.1016/j.compbiomed.2021.104363.

References

[1] T. Thanh Le, Z. Andreadakis, A. Kumar, R. Gómez Román, S. Tollefsen, M. Saville, S. Mayhew, The COVID-19 vaccine development landscape, Nat. Rev. Drug Discov. 19 (2020) 305–306, https://doi.org/10.1038/d41573-020-00073-5.

[4] J. Guarner, Three emerging coronaviruses in two decades, Am. J. Clin. Pathol. 153 (2020) 420–421, https://doi.org/10.1093/ajcp/aqaa029.

[5] M.I. Khan, Z.A. Khan, M.H. Baig, I. Ahmad, A.-E. Farouk, Y.G. Song, J.-J. Dong, Comparative genome analysis of novel coronavirus (SARS-CoV-2) from different geographical locations and the effect of mutations on major target proteins: an in silico insight, PloS One 15 (2020), e0238344, https://doi.org/10.1371/journal.pone.0238344.

[6] M.A. Shereen, S. Khan, A. Kazmi, N. Bashir, R. Siddique, COVID-19 infection: origin, transmission, and characteristics of human coronaviruses, J. Adv. Res. 24 (2020) 91–98, https://doi.org/10.1016/j.jare.2020.03.005.

[7] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-J. Zheng, L. Xu, E.C. Holmes, Y.-Z. Zhang, A new coronavirus associated with human respiratory disease in China, Nature 579 (2020) 265–269, https://doi.org/10.1038/s41586-020-2008-3.

[8] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao, Z.-L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature 579 (2020) 270–273, https://doi.org/10.1038/s41586-020-2012-7.

[9] H. Li, L. Yang, F. Liu, X. Ma, P. He, W. Tang, X. Tong, J. Zuo, Overview of therapeutic drug research for COVID-19 in China, Acta Pharmacol. Sin. 41 (2020) 1133–1140, https://doi.org/10.1038/s41401-020-0438-y.

[10] A. Hegyi, J. Ziebuhr, Conservation of substrate specificities among coronavirus main proteases, J. Gen. Virol. 83 (2002) 595–599, https://doi.org/10.1099/0022-1317-83-3-595.

[11] Z. Jin, Y. Zhao, Y. Sun, B. Zhang, H. Wang, Y. Wu, Y. Zhu, C. Zhu, T. Hu, X. Du, Y. Duan, J. Yu, X. Yang, X. Yang, K. Yang, X. Liu, L.W. Guddat, G. Xiao, L. Zhang, H. Yang, Z. Rao, Structural basis for the inhibition of SARS-CoV-2 main protease by antineoplastic drug carmofur, Nat. Struct. Mol. Biol. 27 (2020) 529–532, https://doi.org/10.1038/s41594-020-0440-6.

[12] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L.W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, H. Yang, Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors, Nature 582 (2020) 289–293, https://doi.org/10.1038/s41586-020-2223-y.

[13] H. Iftikhar, H.N. Ali, S. Farooq, H. Naveed, S. Shahzad-ul-Hussan, Identification of potential inhibitors of three key enzymes of SARS-CoV2 using computational approach, Comput. Biol. Med. 122 (2020) 103848, https://doi.org/10.1016/j.compbiomed.2020.103848.

[14] S. Liu, Q. Zheng, Z. Wang, Potential covalent drugs targeting the main protease of the SARS-CoV-2 coronavirus, Bioinformatics 36 (2020) 3295–3298, https://doi.org/10.1093/bioinformatics/btaa224.

[15] A. Fischer, M. Sellner, S. Neranjan, M. Smieško, M.A. Lill, Potential inhibitors for novel coronavirus protease identified by virtual screening of 606 million compounds, Int. J. Mol. Sci. 21 (2020) 3626, https://doi.org/10.3390/ijms21103626.

[16] M. Rameshrad, M. Ghafoori, A.H. Mohammadpour, M.J.D. Nayeri, H. Hosseinzadeh, A comprehensive review on drug repositioning against coronavirus disease 2019 (COVID19), Naunyn-Schmiedeberg's Arch. Pharmacol. 393 (2020) 1137–1152, https://doi.org/10.1007/s00210-020-01901-6.

[17] D. Rubin, K. Chan-Tack, J. Farley, A. Sherwat, FDA approval of remdesivir — a step in the right direction, N. Engl. J. Med. 383 (2020) 2598–2600, https://doi.org/10.1056/NEJMp2032369.

[18] N. Tsomaia, Peptide therapeutics: targeting the undruggable space, Eur. J. Med. Chem. 94 (2015) 459–470, https://doi.org/10.1016/j.ejmech.2015.01.014.

[19] A.J. Quartararo, Z.P. Gates, B.A. Somsen, N. Hartrampf, X. Ye, A. Shimada, Y. Kajihara, C. Ottmann, B.L. Pentelute, Ultra-large chemical libraries for the discovery of high-affinity peptide binders, Nat. Commun. 11 (2020) 3183, https://doi.org/10.1038/s41467-020-16920-3.

[20] O.N. Silva, O.L. Franco, W.F. Porto, Beta-lactamase inhibitor peptides as the new strategies to overcome bacterial resistance, Drugs Today 54 (2018) 737, https://doi.org/10.1358/dot.2018.54.12.2895652.

[21] P. Karoyan, V. Vieillard, L. Gómez-Morales, E. Odile, A. Guihot, C.-E. Luyt, A. Denis, P. Grondin, O. Lequin, Human ACE2 peptide-mimics block SARS-CoV-2 pulmonary cells infection, Commun. Biol. 4 (2021) 197, https://doi.org/10.1038/s42003-021-01736-8.

[22] K. Kardani, A. Bolhassani, Exploring novel and potent cell penetrating peptides in the proteome of SARS-COV-2 using bioinformatics approaches, PloS One 16 (2021), e0247396, https://doi.org/10.1371/journal.pone.0247396.

[23] D.D. Boehr, K. Draker, K. Koteva, M. Bains, R.E. Hancock, G.D. Wright, Broad-spectrum peptide inhibitors of aminoglycoside antibiotic resistance enzymes, Chem. Biol. 10 (2003) 189–196, https://doi.org/10.1016/S1074-5521(03)00026-7.

[24] R.G. Maroun, S. Gayet, M.S. Benleulmi, H. Porumb, L. Zargarian, H. Merad, H. Leh, J.-F. Mouscadet, F. Troalen, S. Fermandjian, Peptide inhibitors of HIV-1 integrase dissociate the enzyme oligomers †, Biochemistry 40 (2001) 13840–13848.

[25] L. Huang, D.J. Sexton, K. Skogerson, M. Devlin, R. Smith, I. Sanyal, T. Parry, R. Kent, J. Enright, Q. Wu, G. Conley, D. DeOliveira, L. Morganelli, M. Ducar, C. R. Wescott, R.C. Ladner, Novel peptide inhibitors of angiotensin-converting enzyme 2, J. Biol. Chem. 278 (2003) 15532–15540, https://doi.org/10.1074/jbc.M212934200.

[26] A.E. Leffler, A. Kuryatov, H.A. Zebroski, S.R. Powell, P. Filipenko, A.K. Hussein, J. Gorson, A. Heizmann, S. Lyskov, R.W. Tsien, S.F. Poget, A. Nicke, J. Lindstrom, B. Rudy, R. Bonneau, M. Holford, Discovery of peptide ligands through docking and virtual screening at nicotinic acetylcholine receptor homology models, Proc. Natl. Acad. Sci. Unit. States Am. 114 (2017) E8100–E8109, https://doi.org/10.1073/pnas.1703952114.

[27] Z. Wang, S. Zhang, W. Wang, F. Feng, W. Shan, A novel angiotensin I converting enzyme inhibitory peptide from the milk casein: virtual screening and docking studies, Agric. Sci. China 10 (2011) 463–467, https://doi.org/10.1016/S1671-2927(11)60026-6.

[28] S. Ansar, U. Vetrivel, PepVis: an integrated peptide virtual screening pipeline for ensemble and flexible docking protocols, Chem. Biol. Drug Des. 94 (2019) 2041–2050, https://doi.org/10.1111/cbdd.13607.

[29] A. Kumar, P.K. Gupta, A. Srivastava, A review of modern technologies for tackling COVID-19 pandemic, Diabetes Metab. Syndr. Clin. Res. Rev. 14 (2020) 569–573, https://doi.org/10.1016/j.dsx.2020.05.008.

[30] M. Harris, Y. Bhatti, J. Buckley, D. Sharma, Fast and frugal innovations in response to the COVID-19 pandemic, Nat. Med. 26 (2020) 814–817, https://doi.org/10.1038/s41591-020-0889-1.

[31] N. Eroshenko, T. Gill, M.K. Keaveney, G.M. Church, J.M. Trevejo, H. Rajaniemi, Implications of antibody-dependent enhancement of infection for SARS-CoV-2 countermeasures, Nat. Biotechnol. 38 (2020) 789–791, https://doi.org/10.1038/s41587-020-0577-1.

[32] A.M. Arvin, K. Fink, M.A. Schmid, A. Cathcart, R. Spreafico, C. Havenar-Daughton, A. Lanzavecchia, D. Corti, H.W. Virgin, A perspective on potential antibody-dependent enhancement of SARS-CoV-2, Nature 584 (2020) 353–363, https://doi.org/10.1038/s41586-020-2538-8.

[33] A.W.D. Edridge, J. Kaczorowska, A.C.R. Hoste, M. Bakker, M. Klein, K. Loens, M. F. Jebbink, A. Matser, C.M. Kinsella, P. Rueda, M. Ieven, H. Goossens, M. Prins, P. Sastre, M. Deijs, L. van der Hoek, Seasonal coronavirus protective immunity is short-lasting, Nat. Med. (2020), https://doi.org/10.1038/s41591-020-1083-1.

[34] C. Voshavar, Protease inhibitors for the treatment of HIV/AIDS: recent advances and future challenges, Curr. Top. Med. Chem. 19 (2019) 1571–1598, https://doi.org/10.2174/1568026619666190619115243.

[35] M.M. Waldrop, The chips are down for Moore's law, Nature 530 (2016) 144–147, https://doi.org/10.1038/530144a.

[36] T. Joshi, P. Sharma, T. Joshi, H. Pundir, S. Mathpal, S. Chandra, Structure-based screening of novel lichen compounds against SARS Coronavirus main protease (Mpro) as potentials inhibitors of COVID-19, Mol. Divers. (2020), https://doi.org/10.1007/s11030-020-10118-x.

[37] M. Tsuji, Potential anti-SARS-CoV-2 drug candidates identified through virtual screening of the ChEMBL database for compounds that target the main coronavirus protease, FEBS Open Bio 10 (2020) 995–1004, https://doi.org/10.1002/2211-5463.12875.

[38] H. Yang, W. Xie, X. Xue, K. Yang, J. Ma, W. Liang, Q. Zhao, Z. Zhou, D. Pei, J. Ziebuhr, R. Hilgenfeld, K.Y. Yuen, L. Wong, G. Gao, S. Chen, Z. Chen, D. Ma, M. Bartlam, Z. Rao, Design of wide-spectrum inhibitors targeting coronavirus main proteases, PLoS Biol. 3 (2005) e324, https://doi.org/10.1371/journal.pbio.0030324.

[39] V. Roviello, D. Musumeci, A. Mokhir, G.N. Roviello, Evidence of protein binding by a nucleopeptide based on a thymine-decorated L-diaminopropanoic acid through CD and in silico studies, Curr. Med. Chem. 28 (2021), https://doi.org/10.2174/0929867328666210201152326.

[40] K. Rawat, P. Kumari, L. Saha, COVID-19 vaccine: a recent update in pipeline vaccines, their design and development strategies, Eur. J. Pharmacol. 892 (2021) 173751, https://doi.org/10.1016/j.ejphar.2020.173751.

[41] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A. J. Olson, AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility, J. Comput. Chem. 30 (2009) 2785–2791, https://doi.org/10.1002/jcc.21256.

[42] O. Trott, A.J. Olson, AutoDock Vina, Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2009), https://doi.org/10.1002/jcc.21334.

[43] W. Dai, B. Zhang, X.-M. Jiang, H. Su, J. Li, Y. Zhao, X. Xie, Z. Jin, J. Peng, F. Liu, C. Li, Y. Li, F. Bai, H. Wang, X. Cheng, X. Cen, S. Hu, X. Yang, J. Wang, X. Liu, G. Xiao, H. Jiang, Z. Rao, L.-K. Zhang, Y. Xu, H. Yang, H. Liu, Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease, Science 368 (2020) 1331–1335, https://doi.org/10.1126/science.abb4489.

[44] J. Batra, A. Szabó, T.R. Caulfield, A.S. Soares, M. Sahin-Tóth, E.S. Radisky, Long-range electrostatic complementarity governs substrate recognition by human chymotrypsin C, a key regulator of digestive enzyme activation, J. Biol. Chem. 288 (2013) 9848–9859, https://doi.org/10.1074/jbc.M113.457382.

[45] C. Gaboriaud, L. Serre, O. Guy-Crotte, E. Forest, J.-C. Fontecilla-Camps, Crystal structure of human trypsin 1: unexpected phosphorylation of Tyr151, J. Mol. Biol. 259 (1996) 995–1010, https://doi.org/10.1006/jmbi.1996.0376.

[46] W.C. Wimley, S.H. White, Experimentally determined hydrophobicity scale for proteins at membrane interfaces, Nat. Struct. Mol. Biol. 3 (1996) 842–848, https://doi.org/10.1038/nsb1096-842.

[47] R. Bhaskaran, P.K. Ponnuswamy, Positional flexibilities of amino acid residues in globular proteins, Int. J. Pept. Protein Res. 32 (2009) 241–255, https://doi.org/10.1111/j.1399-3011.1988.tb01258.x.

[48] K. Guruprasad, B.V.B. Reddy, M.W. Pandit, Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence, Protein Eng. Des. Sel. 4 (1990) 155–161, https://doi.org/10.1093/protein/4.2.155.

[49] W.F. Porto, I.C.M. Fensterseifer, S.M. Ribeiro, O.L. Franco, Joker, An algorithm to insert patterns into sequences for designing antimicrobial peptides, Biochim. Biophys. Acta Gen. Subj. 1862 (2018), https://doi.org/10.1016/j.bbagen.2018.06.011.

[50] W.F. Porto, L. Irazazabal, E.S.F. Alves, S.M. Ribeiro, C.O. Matos, Á.S. Pires, I.C. M. Fensterseifer, V.J. Miranda, E. Haney, V. Humblot, M.D.T. Torres, R.E. W. Hancock, L.M. Liao, A. Ladram, T.K. Lu, C. de la Fuente-Nunez, O.L. Franco, In silico optimization of a guava antimicrobial peptide enables combinatorial exploration for peptide design, Nat. Commun. 9 (2018) 1490, https://doi.org/10.1038/s41467-018-03746-3.

[51] B. Webb, A. Sali, Protein structure modeling with MODELLER, Methods Mol Biol. 1654 (2017) 39–54, https://doi.org/10.1007/978-1-4939-7231-9_4.