Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Informatics in Medicine Unlocked 29 (2022) 100886

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu

Optimizing peptide inhibitors of SARS-Cov-2 nsp10/nsp16 methyltransferase predicted through molecular simulation and machine learning

John R. Hamre III [a], M. Saleet Jafri [a,b,*]

[a] School of Systems Biology, George Mason University, Fairfax, VA, 22030, USA
[b] Center for Biomedical Engineering and Technology, University of Maryland School of Medicine, Baltimore, MD, 21201, USA

ARTICLE INFO

Keywords:
Covid-19
Anti-microbial peptide
Antiviral peptide
Prediction
Machine learning
Molecular simulation

ABSTRACT

Coronaviruses, including the recent pandemic strain SARS-Cov-2, use a multifunctional 2′-O-methyltransferase (2′-O-MTase) to restrict the host defense mechanism and to methylate RNA. The nonstructural protein 16 2′-O-MTase (nsp16) becomes active when nonstructural protein 10 (nsp10) and nsp16 interact. Novel peptide drugs have shown promise in the treatment of numerous diseases and new research has established that nsp10 derived peptides can disrupt viral methyltransferase activity via interaction of nsp16. This study had the goal of optimizing new analogous nsp10 peptides that have the ability to bind nsp16 with equal to or higher affinity than those naturally occurring. The following research demonstrates that in silico molecular simulations can shed light on peptide structures and predict the potential of new peptides to interrupt methyltransferase activity via the nsp10/nsp16 interface. The simulations suggest that misalignments at residues F68, H80, I81, D94, and Y96 or rotation at H80 abrogate MTase function. We develop a new set of peptides based on conserved regions of the nsp10 protein in the *Coronaviridae* species and test these to known MTase variant values. This results in the prediction that the H80R variant is a solid new candidate for potential new testing. We envision that this new lead is the beginning of a reputable foundation of a new computational method that combats coronaviruses and that is beneficial for new peptide drug development.

1. Introduction

Coronaviruses and other RNA viruses have developed a methyltransferase that promotes viral replication via two main approaches. The first, a self-methylation of the viral RNA is required for viral replication as it disguises the viral RNA to appear as the eukaryotic host RNA [1]. Second, the viral methyltransferases stifle the host-resistance reaction by restraining host interferon creation [2]. SARS (SARS-CoV), Covid-19 (SARS-CoV-2), and other coronavirus analogs have developed two methyltransferases, nsp14 and nsp16, which both require an association with nsp10 to activate [3]. Previous studies have shown that there is similarity of the structures of nsp10/nsp14 and nsp10/nsp16 complexes and in their binding interface [4,5].

There have been numerous efforts to find ways to disrupt SARS-CoV-2 MT activity using small molecules [6–9]. Other SARS-CoV studies have shown that synthetic peptides can interfere in the binding of nsp10 to nsp14 or to nsp16 and can diminish or modulate viral replication [10–14]. In these studies, the investigators uncovered several amino acid sequences of small peptides that bind the methyltransferase and upset the oligomerization of nsp10 to nsp14 or to nsp16. Past mutagenesis studies have shown that there is a commonality in some of the amino acids of nsp10 that associate with nsp14 and nsp16 [15,16]. Moreover, recent findings have shown that some compounds are able to interact with the viral nsp10/nsp16 methyltransferase without affecting human methyltransferases [17]. While these investigations used SARS-CoV, they ought to be similarly appropriate for SARS-CoV-2 as nsp10, nsp14, and nsp16 all have a prominent level of homology between the two viruses. The idea of utilizing a medication to obstruct the connection of two viral proteins in SARS-CoV-2 has been demonstrated successfully by the antiviral Remdesivir, which inhibits the relationship of nsp7 and nsp12 [3].

Peptide drugs have demonstrated effectiveness in many diseases, with more than 60 affirmed and more than 150 in active progress [18,19]. Peptide drugs offer advantages over small molecule drugs and

* Corresponding author. School of Systems Biology, George Mason University, Fairfax, VA, 22030, USA.
E-mail address: sjafri@gmu.edu (M.S. Jafri).

https://doi.org/10.1016/j.imu.2022.100886
Received 23 December 2021; Received in revised form 4 February 2022; Accepted 16 February 2022
Available online 28 February 2022
2352-9148/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

biologics. Small molecule drugs can have many complications and adverse events as their small size permits them to react with many unspecified targets [20]. Biologics can be massive particles, and while more specific than small molecules, these proteins have issues like low bioavailability, penetrability issues, and metabolic unpredictability. Peptide drugs fall between these two limits and can circumvent these difficulties [20].

Antimicrobial peptides (AMP) are one of the first pathways that are upregulated when a host is invaded by microbes. There are numerous AMP classes found in nature that show antimicrobial effects and they have been widely studied, defensins and the human cathelicidin LL-37 have been some of the most studied [21]. AMPs are usually cationic and amphiphilic α-helical peptide molecules, and the mechanism of action for protection is generally thought to be that the cationic properties of AMPs attract and interact with negatively charged microbial cells [22]. Recently it has been found that AMPs can be used for antiviral applications and research and that AMPs with antiviral activity are called antiviral peptides (AVPs) [23]. Initial development for AVP mining research began by using insects because they are rich in AMPs and exist almost everywhere terrestrial [23].

Natural products that already occur in bacteria, viruses, plants and animals have already been successful in the development of new peptide discoveries for disease treatments [24]. Nature has implemented most of the engineering that elucidates a foundational molecule to serve as a design template. These underpinning compounds are consequently modified to attempt to development a different product with higher selectivity, potency, or stability, to be used for therapeutic advancements [25]. More often than not, it is difficult for drug developers to distinguish which modifications will lead to the desired end result, leading to unproductive and costly in vitro or in vivo animal work [26]. However, with the advent of novel, highly accurate computational simulations, many proposed new molecules can be assessed using physics-based methods much more easily.

Machine learning has been applied to features from molecular simulations previously with different goals and relative success [27]. Experimental quantities such as binding affinities are predicted using structural and thermodynamics features. For example, Agrihari and co-workers have PCA to a feature set including phenotype prediction software and other measures from molecular simulations coupled to PCA to make qualitative predictions on severity of phenotype [28]. Sinha and Wang created a feature set including RMSD (Root-mean-square-deviation), RMSF (Root-mean-square-fluctuations), Rg (Radius of gyration), SASA (Solvent accessible surface area), NH bond (hydrogen bond) and Covariance analysis calculated from molecular dynamics simulations for machine learning prediction on whether unclassified variants of the BRCA 1 gene were cancerous or non-cancerous [29]. Gebhard and co-workers predicted solvation energies using machine learning applied to a feature set consisting of intermolecular and intramolecular energies, Lenard-Jones potentials, SASA and Rg [30]. In another study, Kumar and Purohit showed that the RMSD, RMSF, radius of gyration, docking energy, total energy, and protein-solvent interactions were altered in the mutant predicted to be cancer causing [31]. Others have predicted ligand binding efficiency using molecular simulation and machine learning [32,33].

The Molecular Dynamics Phenotype Prediction Model (MDPPM) has proven to be accurate in predicting variant phenotypes and the severity of a disease through structural analysis of a protein over time [34,35]. The model begins by taking trusted three-dimensional protein structures from databanks, simulates the entire structure or portions of it by using molecular dynamics software, then utilizes machine learning to predict what the variant classification or disease-state phenotype will be. The following research utilizes the MDPPM, while developing further an investigational study of natural variants of a nsp10-derived peptide and its relationships to methyltransferase activity.

2. Methods

2.1. Reference protein data bank (PDB) structure and structural variant selection

The reference file used to derive peptide sequences was located at the protein databank (https://www.rcsb.org/retrieved on 12/18/2021), PDBID 6ZPE for SARS CoV-2 nsp10. The molecule was removed of all residues except for the peptide F68 to Y96 used for simulations. Variants were introduced into the new peptide via VMD mutator tool and initial mutations were chosen from the literature where they were shown to cause changes in methyltransferase activity of whole protein interactions of nsp10 and nsp16 [15,36]. These variants were G70A, G94A, G94D, H80A, K93A, K95A, R78A, R78G, S72A, Y96A, Y96F, Y96I, Y96V. New variant sequences were selected due to their occurrence in semi-conserved residues at the sites within the region from F68 to Y96 in the *Coronaviridae* subfamily based upon the multiple sequence alignment of a subsequence of nsp10 from Bouvet et al. [15]. The foundation of the hypothesis that peptides may have an effect on viral replication was that the Y96F variant displayed increased methyltransferase activity in the literature and had occurred in other virus sequences in the subfamily [16]. The new peptide variants chosen were C73I, C73V, C79A, F68Y, H80R, K93R, L92F, L92Y, Y96C, Y96W. The 125 residues shown in the PDB file 6ZPE used are shown below with the residues from F68 to Y96 shown in bold with the substituted residues shown in red.

```
>6ZPE_1|Chain A|Replicase polyprotein 1ab|Severe acute respiratory syndrome
coronavirus 2 (2697049)

1       TMGNST VLSFCAFAVD AAKAYKDYLA SGGQPITNCV KMLCTHTGTG 50
51  QAITVTPEAN MDQESFGGAS CCLYCRCHID HPNPKGFCDL KGKYVQIPTT 100
101 CANDPVGFTL KNTVCTVCGM WKGYGCSCD                       129
```

2.2. Molecular dynamics simulation methods

Molecular dynamics simulations were deployed by using Nanoscale Molecular Dynamics (NAMD) using the CHARMM36 forcefield [36–38]. Three REMD simulations were carried out for each peptide variant at 308K–312K due to the instability of a few of the variant peptides. We found that at higher temperatures the stabilities varied greatly, and some structures fell apart (data not shown). Replica exchange molecular dynamics (REMD) simulations were employed to sample the conformational ensembles of the nsp10 variants and wild-type proteins at the temperature range corresponding to normal human body temperatures. To verify that replicas performed a random walk across the temperature range, albeit small, we computed the acceptance rate of the fraction of successful replica exchanges and found it to be greater than 95% when averaged over all REMD temperatures. We decided to use this temperature range because we did not need to facilitate any higher temperatures. This was due to the fact that the protein was already folded and because we chiefly wanted to test the peptide variant stabilities and smaller shifts in structure that occur close to body temperature. All simulations were carried out for more than 20 ns as the peptide was taken from the larger protein and therefore folding assessments with

longer simulation times were not necessary. The General Born Implicit Solvent GBIS model was used for all simulations with a 2 fs timestep. ShakeH algorithm was set to "on," salt concentrations were set at 0.3 M, and the surface tension was set to 0.006 kcal/mol/Å, with Langevin dynamics used for temperature control. Scripts and starting PDB structures for the molecular dynamics simulations are archived in the George Mason University Dataverse at https://doi.org/10.13021/orc2020/OW4T3K (retrieved on 1/26/2022).

2.3. Feature extraction

Utilizing VMD, the phi and psi dihedral angles of the protein were extracted for the desired timepoints to be used as the initial feature set matrix for the analytical calculations. A set of raw data consisting of the dihedral angle for each variant was copied into a matrix and utilized for all MDPPM computations. For at least two of the simulations, at least 500 of the last PDB file frames were placed into the matrix and used for machine learning or PCA analysis. Using the last 500 frames allows for two measures to strengthen the analysis: 1) the root mean square deviation smooths out in the trajectory, which indicate that the peptide has stabilized, and 2) the number of frames is enough for vigorous statistical analyses. The average peptide structures of the simulation trajectories were found by averaging the frames by using a small TCL code. Average structures were viewed qualitatively for fit and structure changes and atomic distance measurements were viewed or taken via PyMol (Schrodinger, LLC, New York, NY). The TCL code for extracting the phi and psi can be found at the George Mason University Dataverse at https://doi.org/10.13021/orc2020/OW4T3K (retrieved on 1/26/2022).

As the feature set is large, the data dimensionality was reduced using a principal component analysis (PCA) using RStudio (RStudio Team, 2021). The PCA applies a linear transformation (change of coordinate systems) to rank the new axes or principal components (PCs) in order of PCs representing most of the spread of the data to the least spread of the data. The PCs were used to find the Euclidean distance measurements on the resulting data. Euclidean distances were calculated by measuring from centroid to centroid from the wild type to the variants.

In order to fit the peptide variant data to methyltransferase activity, we gathered nsp10/nsp16 interaction and methyltransferase data. The empirical data was taken from Bouvet et al. (2014) and is gathered by measuring energy transfer as % BRET in in vitro interaction assays, which is used as a proximity-based assay to monitor protein–protein interactions and conformational rearrangements in live cells, as well as 2′-O-MTase activity [39]. The Euclidean distances were used to predict the experimental values by fitting a non-linear curve, $y = a(1\text{-}e^{-bx})+c$, where x is the distance and y is the measured value according to the MDPPM method [34,35]. The predicted and experimental methyltransferase data were correlated by using a Pearson's statistical test. This involves calculating the Pearson Correlation Coefficient and then using a two-sided p-value for the t-distribution with n-2 degrees of freedom to test for significance $t = r\sqrt{n-2}/\sqrt{1-r^2}$. The predicted values calculated from the simulations were correlated to these data for comparisons to methyltransferase activity and the structure of the peptide. We correlated our predictions measurements to % BRET, % Interaction between nsp10 and nsp16 and & 2′-O-MTase Activity (referred hereafter in this study as MTase for methyltransferase).

To predict if a new variant would have the same, or possibly increased, MTase activity, we used the random forest algorithm. For theses predictive analyses, the data matrix is processed further by a subsequent PCA analysis on the PCs themselves, which serves to reduce the dimensions even further. This matrix is used for the predictive models where we use random forest to train and predict. For each variant tested, we remove its data before training, train on all remaining data points or specific columns of PCs, then add in the variant to be tested and the results of the random forest are either an H or a L, as per Table 2. Accuracy of the PCs for prediction can be assessed beforehand with a random forest which results in graphs similar to Fig. 6.

Initial tests used only the variants that had in vivo or in vitro data to qualify and verify method success, which was a strong correlation to previous data from the PCA analysis. In order to reduce bias in the training data and to mimic clinical settings for unknown variants, we pulled the variant to be tested out, trained on the remaining variants, and the new variant was added to the already trained data to assess predicted classification Based upon the available MTase data, we classified any variants with lower than <90% as having "low" activity and >90% as "high." When we were satisfied with the prediction model on the variants with data, we began by adding in the newly derived peptide variants from the natural product research. Only those new variants classified as "high" were categorized as potential new candidates for future research. The model was developed to take a false positive approach, where the variant was deemed more likely to cause a decrease in MTase activity, so that we errored on the side of caution when screening new variants. The initial code for the MDPPM used in previous studies can be found at https://github.com/MDPPM/initialCode (retrieved on 12/18/2021). The codes used for our analysis are available at https://doi.org/10.13021/orc2020/OW4T3K (retrieved on 1/26/2022).

3. Results

The first step in predictive modeling and assessment using the MDPPM is an all-atom simulation of the protein followed by a PCA analysis on the phi and psi dihedral angles to reduce the dimensionality of the data [34,35]. Euclidean distances between the variants and the wild type are then calculated using the PCs. Table 1 shows the distances calculated from the phi an psi dihedral angles and correlates those to the Bouvet et al. (2014) data. The Euclidian distances were then mapped to the experimental data using non-linear regression to the formula $y = a(1\text{-}e^{-bx})+c$, where x is the distance and y is the measured value in according with the MDPPM method (Table 1 and Fig. 1). These data show that our model results strongly correlate to all three endpoints with statistically significant P-values using Pearson correlations. Fig. 1(a) shows the predicted values (red) and the experimental values (blue) for % BRET. Fig. 1(b) shows the correlations between the experimental and observed which are fit with a linear regression line with slope 1 passing through the origin. The plots for the % Interaction and % MTase activity are shown in Fig. 1(c–f). The correlation coefficient for these fits were 0.88, 0.77, and 0.80. The Pearson correlations we 0.82, 0.73, and 0.81 for % BRET, % Interaction and for % MTase activity, respectively These Pearson correlations coefficients were statistically significant with P-value of 0.000357, 0.002173, and 0.000249 for % BRET, % Interaction and for % MTase activity, respectively using two-sided p-value for the t-distribution with n-2° of freedom where n is the number of variants.

Next, we examined the average structures of each peptide throughout the simulation. When comparing with the current literature, there are conserved regions in the region of nsp10 where we extracted the peptides and designed the experiment to first compare the conserved regions of the wild-type peptide to the variants [16]. Initially, we performed a comparative analysis of the peptides using a general alignment in PyMol. We began by qualitatively examining the structural differences of the peptide variants as compared to the wild-type peptide so as to gauge structural dissimilarities and similarities. Notably, variants G94D and Y96F modulated the MTase activity to be very low and very high, respectively. G94D is one of only two variant peptides from the literature that resulted in zero methyltransferase. Fig. 2 shows that the G94D average structure did not fit well onto the wild-type protein. Most notably we see that there are misalignments at residues F68, H80, I81, D94, and Y96. However, the Y96F variant has an almost identical structural alignment.

We assessed each variant to compare what the overlays revealed as

Table 1
Comparison of predicted values to experimental data.

Variant	Experimental			Predicted			
	% BRET	% Interaction	% 2′-0- MTase Activity	Euclidean Distance	% BRET	% Interaction	% 2′-0- MTase Activity
WT	100	100	100	0	112.40	111.90	132.70
G70A	31 ± 7.4	50	32 ± 0.8	5.17	24.43	25.50	11.61
S72A	60 ± 8.6	6	22 ± 1.1	4.61	28.170	30.42	16.90
R78A	9 ± 4.5	10	2 ± 0.1	3.3	40.39	44.99	34.18
R78G	35 ± 1.8	6	9 ± 0.3	3.64	36.66	40.74	28.95
H80A	ND	78	60 ± 3.1	3.51	38.03	42.32	30.88
K93A	35 ± 2.2	54	9 ± 1.2	3.58	37.29	41.47	29.83
K93E	7 ± 5.8	16	0 ± 9	4.38	29.93	32.64	19.45
G94A	59 ± 2.4	80	85 ± 1.2	3.19	41.69	46.45	36.00
G94D	16 ± 4.9	7	0 ± 0.1	5.76	21.21	21.02	6.97
K95A	80 ± 3.9	83	71 ± 2.5	3.35	39.81	44.35	33.37
Y96A	30 ± 6,2	6	15 ± 0.6	3.84	34.67	38.40	26.15
Y96F	123 ± 18	124	163 ± 10	0.1	108.79	108.96	127.84
Y961	12 ± 2.3	3	4 ± 0.4	6.47	18.11	16.43	2.49
Y96V	20 ± 3.4	0	5 ± 0.1	2.97	44.45	49.48	39.86

	% BRET	% Interaction	% 2′-0- MTase Activity
Pearson Correlation	0.82	0.73	0.81
R^2	0.88	0.77	0.80
Root Mean Squared Error (RMSE)	394.98	640.51	655.64

Table 1. Correlation of the predicted values from the PCA analysis as compared to experimental nsp10/nsp16 data. The Pearson correlations were statistically significant with P-value of 0.000357, 0.002173, and 0.000249 for % BRET, % Interaction and for % MTase activity, respectively using two-sided p-value for the t-distribution with n-2° of freedom where n is the number of variants.

Table 2
Prediction of MTase activity of test variants.

Variant	Predicted	Actual	Test Variant	Prediction
WT	H	H	F68Y	L
G70A	L	L	C73I	L
S72A	L	L	C73V	L
K78A	L	L	C79A	L
K78G	L	L	H80R	H
H80A	L	L	L92F	L
K93A	L	L	L92Y	L
K93E	L	L	K93R	L
G94A	L	L	Y96C	L
G94D	L	L	Y96W	L
K95A	L	L		
Y96A	L	L		
Y96F	H	H		
Y96I	L	L		
Y96V	L	L		

Table 2. Left, prediction using random forest of variants with known MTase activity in whole protein data designated as having L (low) or H (high) MTase activity. Sensitivity, specificity, accuracy for positive (H) versus negative (L) were all 100%. The right portion of the table shows the new variants derived from other coronavirus sequences and their prediction of a future candidate for testing "H" or not at candidate "L." A value of H predicts the MTase value to be more than 90%.

compared to methyltransferase activity. We began by viewing the overlays of G94D and K93E, both having complete zero methyltransferase activity, and used a space filling model to identify any residues that changed in both peptide simulations. It was discovered that the H80 residue in both G94D and K93E shifted clockwise approximately 90° to an almost identical position in each variant (Fig. 3). Next, we looked at the other H80 residues in each variant. In all cases except for the Y96"X" variants, the H80 residues displayed the same rotation of H80 into a similar position as the G94D and K93E variants. Finally, for the H80A variant, because it replaced the histidine with an alanine, we explored the alanine positioning. The alanine residue also moved but only rotated approximately 45° as opposed to the other H80 rotamers. This H80 rotation was detected to also causes a knock-on effect that moves the, H80, I81, and D82 residues further to the left as compared to the wild type (Fig. 4). We calculated this effect to produce ~2.5 Å alteration of these residues in the variants R78A and R78G. As compared to histidine, alanine does not have a positive charge and it is hydrophobic, but it too caused the knock-on shift. Therefore, we reasoned, that because the variants that caused this shift had middle to low MTase values, we ruled out any peptide containing this phenotype as contenders for future testing as they would possibly affect methyltransferase activity.

In the variants containing a Y96"X" change, Y96V, Y96I, Y96A, T96F, the H80 residue did not shift. Consequently, we sought out other structural changes that may have negatively affected the methyltransferase activity in these variants. In Fig. 5 we compare the Y96"X" variants to the interaction regions of nsp10 to nsp14, and to nsp16. We observed that all Y96"X" variants displayed an S72 residue that protrudes further than wild type, including Y96F. For the Y96F variant, that was the only notable change except for the F96 residue itself. In the other variants, Y96V, Y96I, and Y96A, the A71 residue changed, also protruding out further. But again, the most notable change, is the change in Y96 to V, I or A, itself. In all cases, the R group of the Y96 molecule loses a great deal of mass. For Y96F, it should be noted that this variant is also found in many other coronavirus sequences and was the only variant with increased MTase activity [15].

Structurally, for the new variants, they are all mostly unremarkable. We found that none had the H80 rotation, and none appeared to shift the I81 residue. For the Y96"X" variants, Y96C and Y96W, again unremarkable, other than the Y96 residue. Y96C is quite a bit smaller than Y96, and Y96W, as per the R group is larger than Y96.

The random forest algorithm performed on the phi and psi dihedral angles revealed that the top 20 primary torsion predictors of accuracy for classification were residues located in the 70's and 90's portion of the peptide, with psi89 being the exception, situated adjacent to the highly conserved C90 residue (Fig. 6). The C90 amino acid is also involved in zinc chelation and the psi90 angle is the third most accurate from the algorithm. Two other residues involved in zinc chelation were also in the top 15 angles for accuracy, psi74 and psi77. Phi82, located next to the fourth chelating residue H83, was found on the Gini index.

When using random forest for the prediction of MTase activity using the MDPPM, it predicted with 80% accuracy using the leave-one-out strategy and designating three separate classes as "low," "middle," or "high," according to MTase activity (Table 2). When using binomial "high" or "low" classification predictions, the sensitivity was 100%, the

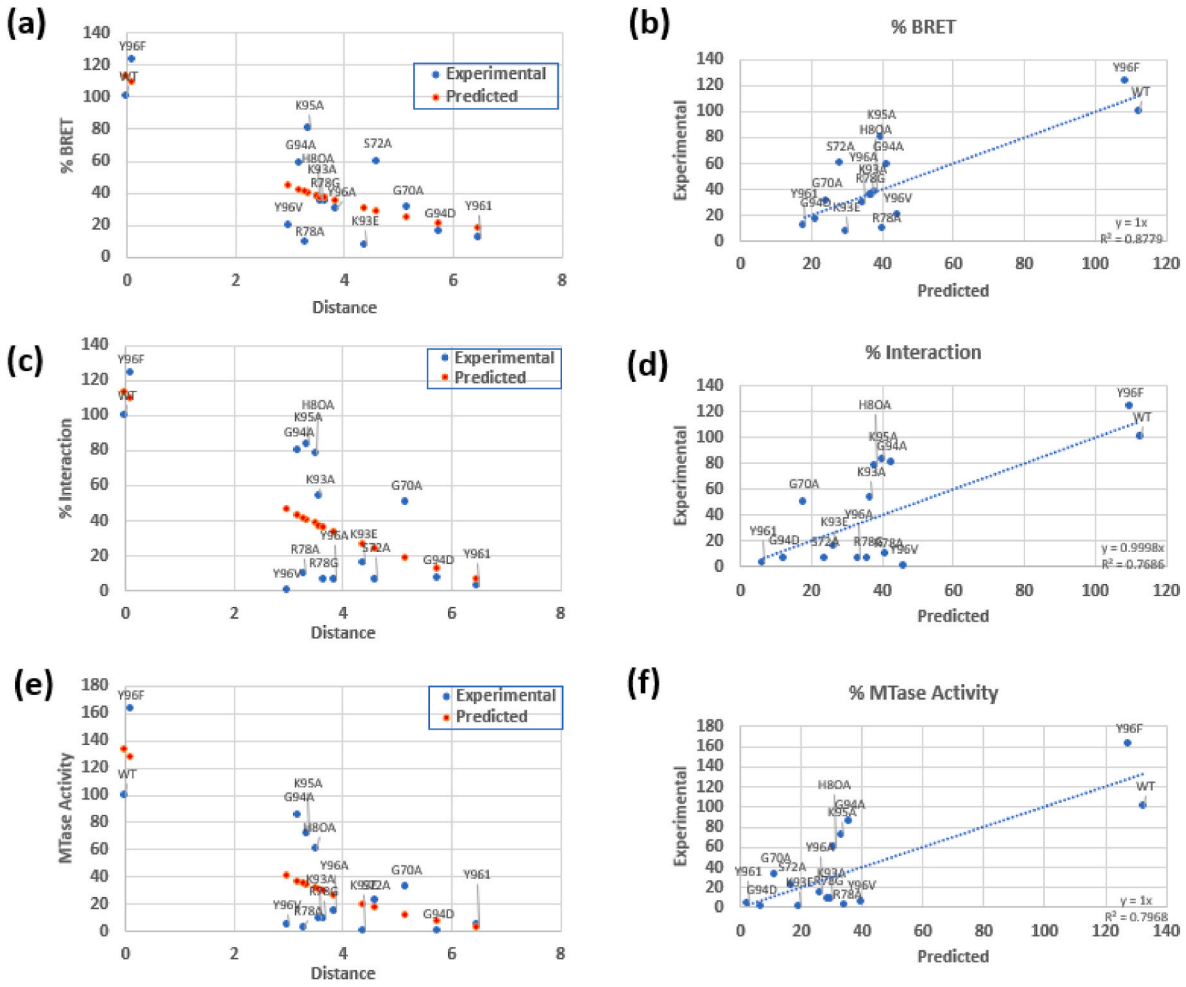

Fig. 1. (a) Experimental (blue) and predicted (red) % BRET values as a function of distance in the PCA space. The predicted values follow the equation (% BRET) = -105.3*(1-$e^{(-0.349*(Distance))}$)+112.4. (b) Experimental vs predicted % BRET values show a high degree of correlation R^2 = 0.8779. (c) Experimental (blue) and predicted (red) % Interaction values as a function of distance in the PCA space. The predicted values follow the equation (% Interaction) = -142*(1-$e^{(-0.214*(Distance))}$)+112.8. (d) Experimental vs predicted % Interaction values show a high degree of correlation R^2 = 0.8779. (e) Experimental (blue) and predicted (red) % MTase Activity values as a function of distance in the PCA space. The predicted values follow the equation (% MTase Activity) = -146.8*(1-$e^{(-0.337*(Distance))}$)+132.7. (f) Experimental vs predicted % MTase Activity values show a high degree of correlation R^2 = 0.8779. Predicted values were obtained using non-linear curve fitting for the experimental values as a function of distance using the web site https://www.colby.edu/chemistry/PChem/scripts/lsfitpl.html (retrieved on 12/7/2021). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

specificity 67%, and the accuracy 93%. In this case, the variant was classified as "low" if the variant decreases the MTase activity lower than 80%.

In the PCA analysis, on the variants with empirical data, we were able to visualize the 3D components of the system and their distances from each other (Fig. 7). The WT, Y96F, and G94A were designated as having "high" MTase activity and we classified all other variants as "low." For the newly proposed variants, we performed an additional PCA analysis, one that included all variants and was plotted for visual analysis. We again used the random forest algorithm to find that PC1 and PC5 clustered the most accurately, and that the new variants were typically clustered away from the other variants in their own PC space (Fig. 8).

4. Discussion

The emergence of new computational methods for drug development have begun to generate potential therapeutic compound structures at an unprecedented rate. One of the largest remaining challenges in peptide development is finding worthwhile new candidates amongst the sheer number of molecular possibilities. Currently, there are a few canonical approaches into how new drugs are designed and sifted through using computation methods and software. Molecular modeling, structure-based drug design, structure-based virtual screening, ligand-based modeling, and molecular dynamics are all used to help determine the relationship between the ligand and the target [40]. Here, we have taken a multi-faceted approach by combining the molecular modeling method with structure-based drug design, while using molecular dynamics to assess the structure in solvent. These all lead to machine learning data analytics for predictive measures.

Antimicrobial peptides have been identified to be widespread in nature as part of the innate immunity in many organisms [41]. Many have been found in nature, however, their use as therapeutic agents raised challenges that include issues with stability, bioavailability, and toxicity. This has resulted in the desire to design or engineer antimicrobial peptides. The methods for designing antimicrobial peptides includes de novo synthesis, chemical modification, site direct mutation, template based design and computer-based design [41]. The computer-based design methods include the use of molecular modeling and machine learning approaches. There have been several studies that have used machine learning via neural networks to design antimicrobial peptides [42–44]. Other efforts have used a variational autoencoder

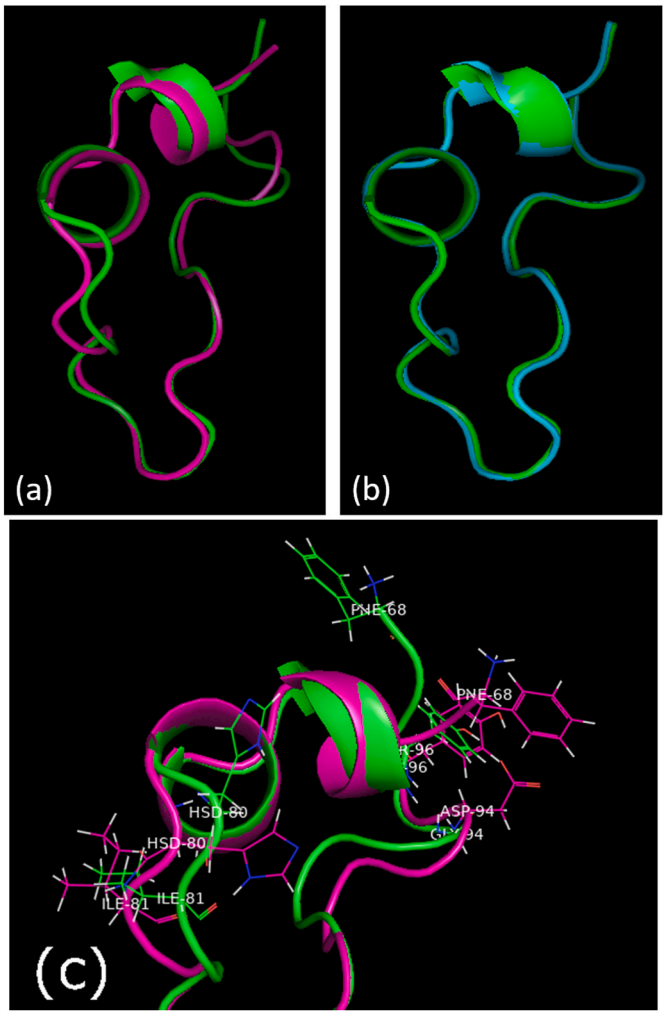

Fig. 2. (a) Wild type (green) and G94D (magenta) overlay displaying large differences in several positions, most notably F68, H80, I81. (b) Wild type and Y96F (cyan) overlay exhibiting nearly complete structure homogeneity with a few subtle differences. (c) enlarged view with notable positions F68, H80, I81, G94D, and Y96 shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

framework to generate and predict the activity of antimicrobial peptides [45]. These studies relied on large databases of antimicrobial peptides to generate the training and test sets [45]. The method presented here differs from these previous studies by focusing on the optimization of antimicrobial peptides by amino acid substitutions.

The MDPPM method is developed to predict the severity of change with missense mutations (changes in single amino acids) [34,35]. While the method is quite powerful there are some issues that need to be addressed in the future. The peptides being compared need to have a large degree of similarity to allow for comparison. The method has not yet been applied to other types of mutations such as deletions, insertions, truncations, and frame shift mutations. Also, due to the reliance on molecular dynamics simulation, the method is computationally expensive. These simulations have been using the implicit solvent model to increase efficiency. In the future we plan to address ways of improving computational efficiency by trying newly developed algorithms for molecular simulation.

The main advantage the proposed method has over existing methods is the high degree of accuracy and the ability to succeed with small data sets. It is geared for refining peptide candidates. Dean and co-workers recently developed a computational framework called PepVAE for the prediction of the activity of antimicrobial peptides [45]. It consists of a variational autoencoder to design the peptide and antimicrobial activity prediction models that use a convolutional neural network that yield prediction with a correlation coefficient of 0.67–0.73. Mulligan has presented a review discussion on the role of computational design in peptide drug discovery [46]. These computational approaches are used to predict drug specificity, solubility and aggregation propensity, permeability, and toxicity [47–54]. These approaches use QSAR method, machine learning, and multiscale modeling. The existing QSAR and machine learning methods train on large databases. Molecular dynamics simulation methods have been used to study smaller lists by

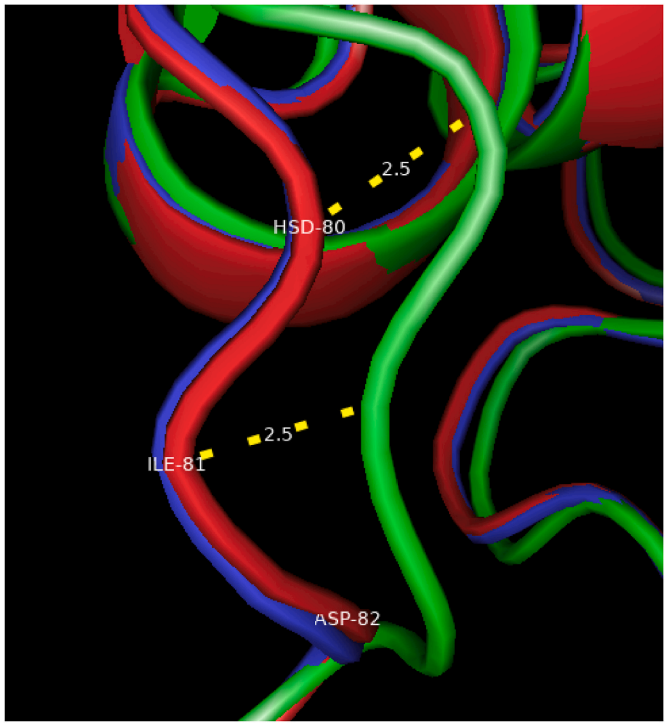

Fig. 4. In the variants R78A (blue) and R78G (tv red) we observe that the 3 residues H80, I81 and D82 are shifted, with the furthest distance being 2.5 Å. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

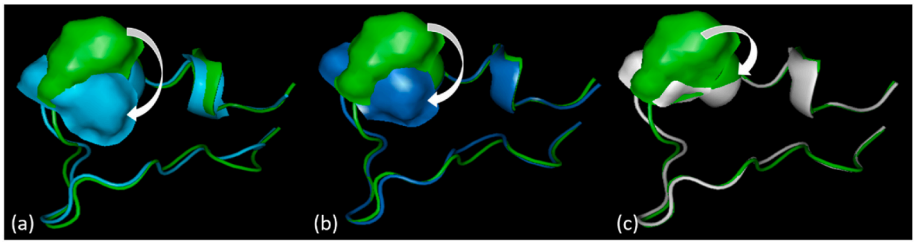

Fig. 3. (a) Histidine rotamer in G94D.(b) K93E, and (c) H80A turning approximately 90°.

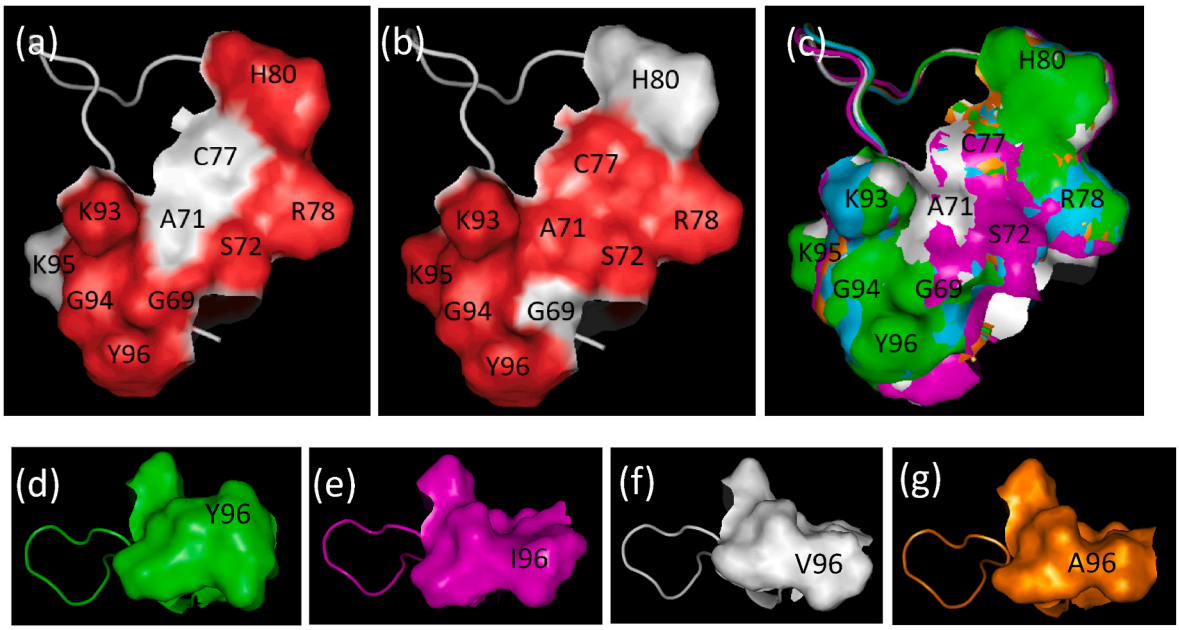

Fig. 5. (a) WT peptide showing residues that interact with nsp14 in red (b) WT residues that interact with nsp16 shown in red. (c.) Y96V white, Y96I magenta, Y96A orange, Y95F cyan, and WT green overlay showing the differences in the variants. Notable qualitative changes include the A71 residue protruding out further than the wild type in all variants, including Y96F. For the Y96F variant, this was the only notable change except for the F96 residue itself. In the other variants Y96V, Y96I, and Y96A, the A71 residue did change, protruding out further. (d–g) Most notable is the change in Y96 to V, I or A, itself. In all cases the molecule loses a great deal of mass, and the protrusion that is seen in the wild type contracts drastically. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

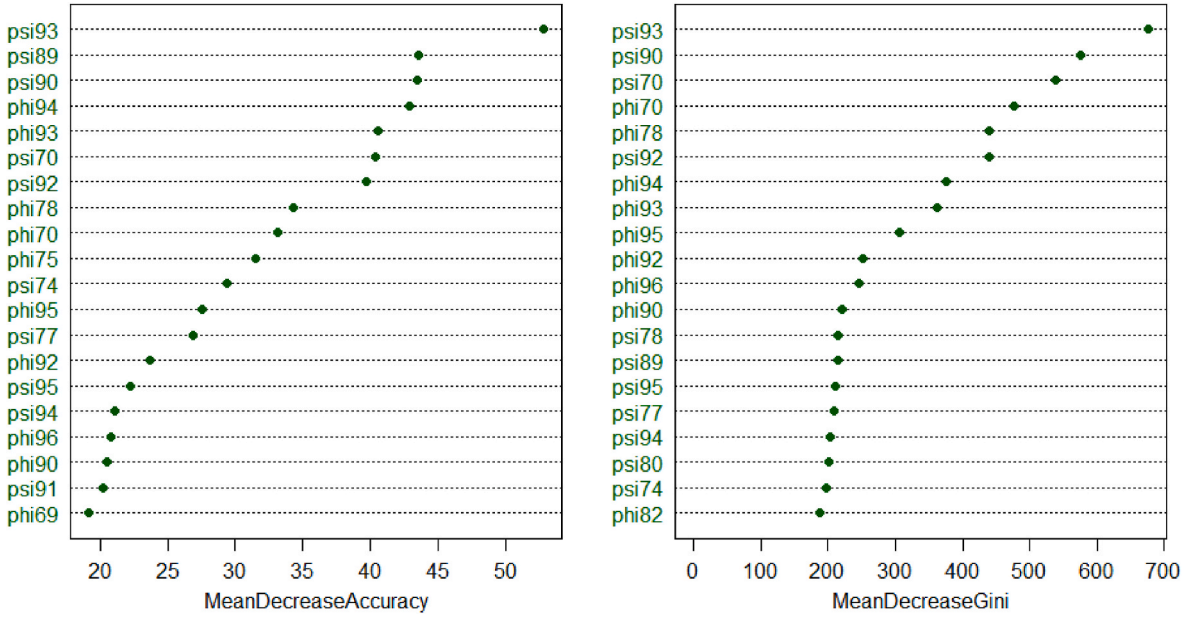

Fig. 6. Random forest output of the top dihedral angles in the peptides. Of note, most of the amino acid residues listed are in the 90's and 70's, with, C74, C77, and C90 involved in the chelation of zinc.

simulating the candidate molecule and its surroundings to calculate physical properties from first principles. These quantities are then used in another model to predict the best candidate peptides. The MDPPM method presented here avoids this additional step by directly using properties of the structural ensemble to predict the peptide efficacy, in this case the effect on MTase activity. Molecular docking simulations of peptide protein binding such as ZDOCK 3.0.2, FRODOCK 2.0, Hex 8.0.0, PatchDock 1.0, ATTRACT and pepATTRACT have also been used in peptide drug development. However in a recent test of accuracy compared to the observed experimental peptide-protein docking the best was only 1.5–63.90% accurate [55]. Docking simulations face the challenges of finding the correct binding site and then orienting the peptide so that it binds correctly to its target. The MDPPM method proposed here avoids these complications by simulating the peptide alone in solution and using the structural ensemble which captures the structural dynamics to predict efficacy. Other approaches include phenotype prediction software where a sequence and variants are submitted to a web server and predictions of whether the variant is benign

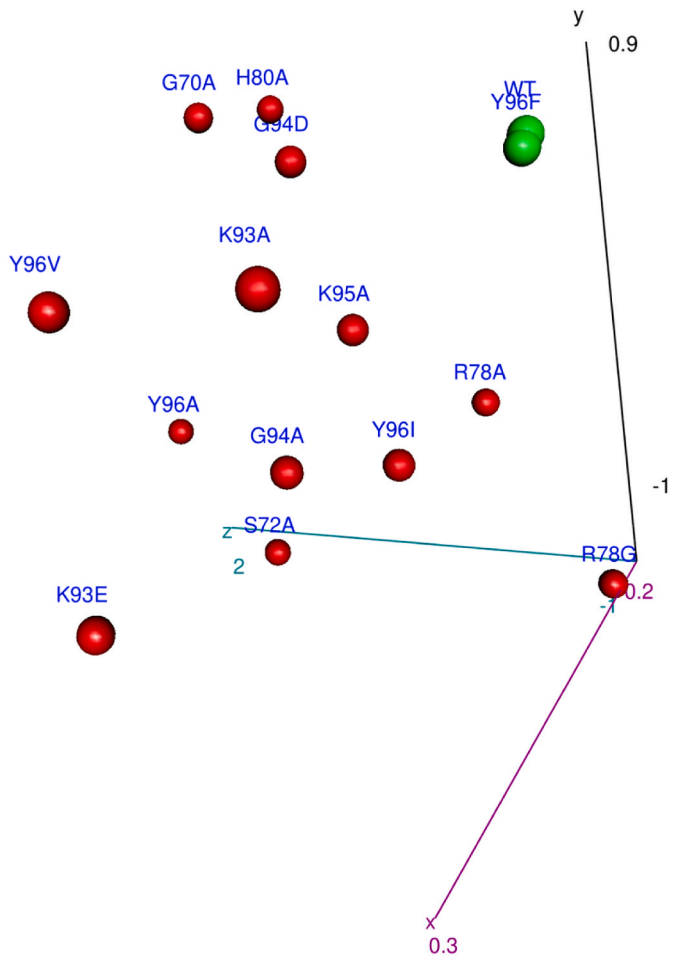

Fig. 7. 3D image of PCA space of the original variants with known empirical data from Bouvet et al. (2014). The WT and Y96F were designated as having "high" MTase activity, >90% (green). We classified all other variants as "low" (red). Low value variants clustered together, and we can see qualitatively that the Y96F and WT distances were very close. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

but offer much lower accuracy than the MDPPM method upon which the method in this paper is based [34,35].

Other computational studies relating to peptide development for SARS-Cov-2 have been performed, but they differ greatly in their scope and success. Can et al. only go so far as to say that nsp10 epitopes will be good candidates for vaccines [56]. Sk et al. investigated the nsp10/nsp16 interface, but only recommend residues that might be investigated in the future such as G69, A71, G70, R78, Y96, three of those were used in this research [57]. Others have done this as well, only venturing to propose hot spots [58]. Ling and Sitthiyotha developed peptide models for SARS-Cov-2, but only were able to suggest new molecules based on molecular docking models [59,60]. Furthermore, these models were not qualified or evaluated against wet lab data. It is well known that docking programs are notoriously inaccurate, with the best scoring at 50.0–60.0% accuracies for pose predictions, and with relatively weak correlations between docking scores and observed binding affinities [61]. This study's advantage was that the peptide was already in the proper pose and that the binding energies in our model are predicted to be close to wild-type energies because they already exist in nature.

Two of the main challenges currently in the development of antiviral and antimicrobial drugs include the emergence of multidrug resistance and the transmission of drug-resistant strains from patient to patient that limit the clinical efficacy of current therapy [62]. For emerging peptide drugs, one of the largest problems is the increased proteolytic instability as compared to not small molecules and monoclonal antibody therapeutics [63]. The advantages to this model are that the stability can be accounted for in the simulations and that the epitope of the peptide is conserved. Further, once peptides are elucidated from the MDPPM, experimental testing using wet lab techniques such as the serum stability assay could be beneficial, with future correlations to peptide stability to the MDPPM to be measured in future studies for machine learning model refinement [64].

Recent epidemiological data has shown that many people are still contracting SARS-CoV-2. There is, therefore, the need to have effective treatments for people infected with SARS-CoV-2. Current treatments of SARS-CoV-2 include two prominent drugs that were used for other indications, hydroxychloroquine (HCQ) and Remdesivir (RDV) [65]. HCQ has had mixed results in Covid19 treatment, however, when coupled with azithromycin it resulted in a good clinical outcome and a virological cure in 973 out of 1061 patients within 10 days (91.7%) [66]. Notably, of those, 9 developed QTc prolongation but did not die of cardiotoxicity. Rare cardiotoxicity adverse events are a large concern among the public. RDV in a compassionate use clinical study showed that there were clinical improvements in 67.9% of patients but that 60.3% had adverse events that included enzyme elevation and renal impairment [67]. Both drugs demonstrate the need for alternative treatments due to safety measures and uncertain effectiveness, hence the need for nsp10/nsp16 conserved region interface drugs that target the virus machinery with higher safety and efficacy profiles.

Targeting MTase activity is a mechanism that can help suppress viral replication. Normally, the host cell will degrade non-human RNA, but in order to overcome this SARS-CoV-2 has two methyltransferases that methylate viral RNA making it appear human to the host cell. The inhibition of methyltransferase has the advantage of allowing the host cell to degrade viral RNA preventing viral replication. The peptide would be specific to the viral methyltransferase and not affect the host, thereby having limited side effects. Lin et al. compared the nsp10/nsp14 structure with the nsp10/nsp16 structure and found them highly similar [4]. Kozielski et al. observed that the binding interface of nsp10 with nsp14 and nsp16 were conserved by screening binding of nsp10 fragments (PDB ID: 7ORR, 7ORU, and 7ORV) [5]. Krafcikova et all compared the structures of the nsp10/nsp16 protein complex and found that while the active site of SARS-CoV-1 and SARS-CoV were conserved, they were not similar to Zika Virus [68].

Experimental studies show that a substitution inactivating the N7-MTase activity substantially attenuated replication, while complete knockout of 2′*O*-MTase knockouts was less effective [69,70]. Based upon these observations, DeCroly et al. has suggested that compounds specifically inhibiting cap-methylating enzymes, either N7-MTase or 2′*O*-MTase or both, could act as potent antiviral agents [71]. Wang et al. has shown that this approach can work in their experiments [11]. Hsu et al. found that nsp14 shuts down host protein synthesis in order to help SARS-CoV-2 subvert protein synthesis for viral proteins [14]. The exoribonuclease activity of nsp14 combines with the MTase activity of nsp7 to promote viral replication. They showed that mutations in the sites of exonuclease or N7-MTase activity of nsp14 abolish its translation inhibition activity. Furthermore, this mechanism is conserved in human coronaviruses.

Several studies have searched for small molecules that can inhibit MTase activity. Bobrovs et al. computationally screened the docking of 7 million compounds with nsp14 and nsp16 and found 80 candidates, 39 for nsp14 and 41 for nsp16 [6]. They experimentally assayed them and found 9 that displayed MTase inhibition with an $IC_{50} < 200$ μM. However, challenges still remain because most of the compounds had poor selectivity for a specific MTase, no cytotoxic effects, and poor cell permeability. Basu et al. computationally screened 5000 compounds and found 4 compounds that were potential inhibitors of nsp14. All 4 displayed anti-viral properties in infected cells and 3 showed synergistic

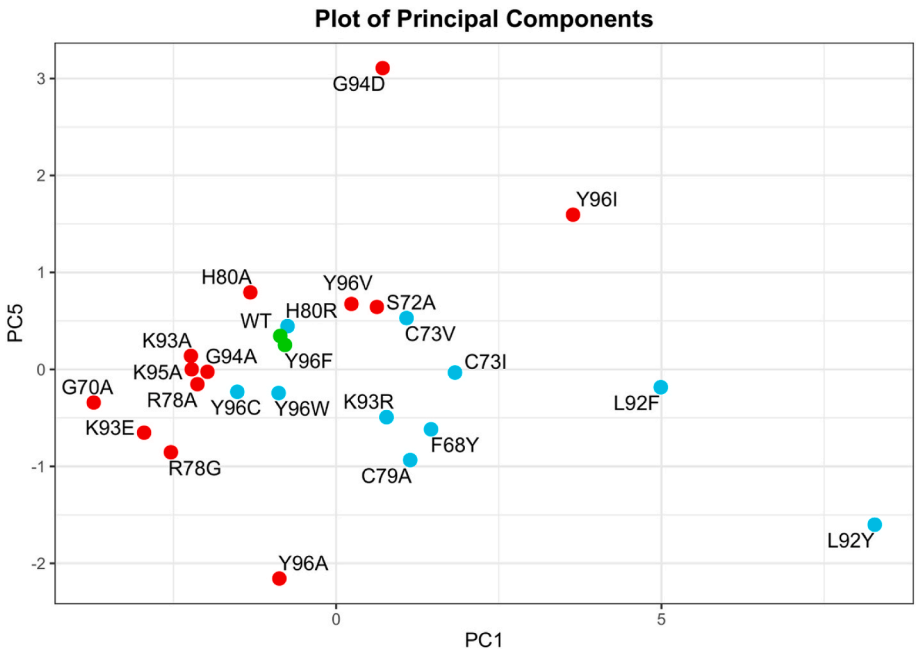

Fig. 8. New PCA space including all new variants of PC1 vs PC5. This is the 2D PCA plot that shows a new PCA analysis with all the variants, including the new variants to be tested (cyan). Variants with high MTase are in green and those in red have low MTase activity. New variants tended to be clustered lowered and to the right. New variants closest to the green cluster, Y96C and H80 were predicted by random forest to be candidates for future testing. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

effects when paired with remdesivir [9]. Bovieva et al. used S-adenosylmethionine derivatives to inhibit the MTase activity of nsp16. None of these molecules were found to be specific to the nsp16 MTase. Despite inhibition of the human MTases no cytotoxicity was observed [7]. Devkova et al. tested 161 in-house synthesized *S*-adenosylmethionine (SAM) competitive MTase inhibitors and SAM analogs and found that 6 that were inhibitors of nsp14. Of these one was found to have an IC_{50} = 70 nM and was selective against 70 human lysine MTases [8].

On the other hand, discovering peptide inhibitors of the interaction between nsp10 and nsp16 to disrupt MTase activity to suppress viral replication has also been an area of active research. Previous SARS-CoV studies have shown that synthetic peptides can interfere in the binding of nsp10 to nsp14 or to nsp16 and can diminish or modulate viral replication [10–13]. Ke et al. showed that two small peptides from the range of amino acids 65–107 of nsp10 successfully inhibited 2′-O-methyltransferase activity of SARS-CoV nsp16/10 complex [13]. Wang et al. designed a peptide called TP29 that was extracted from the nsp10 interaction interface of mouse hepatitis virus (MHV) nsp10 and showed that the peptide inhibits the 2′-O-MTase activity of different coronaviruses in both biochemical assays and in viral replication studies in MHV infection and SARS-CoV replicon models [11]. The % BRET, % Interaction, and % MTase activity data for training the model was gathered from the experimental work of Bouvet et al. [15]. The work presented here, builds upon these studies by using machine learning to predict another candidate peptide that should be studied in experiments to verify efficacy at disrupting MTase activity. These studies differ from the computational work of Dutta et al. who used molecular dynamics simulation of nsp10/nsp16 and that two fragments of nsp16 that bound to nsp10 in their simulation studies [72]. While based on the same idea of disrupting nsp10/nsp16 oligomerization to inhibit MTase activity, they performed computationally expensive studies of peptide fragments of nsp16 binding to the nsp10 protein. However, their approach was not validated with known inhibitory peptides. Our approach on the other hand is much more computationally efficient and used a machine learning model trained and tested on nsp10 peptides measured to affect MTase activity.

Here, we found one candidate, H80R, that is almost 100% likely to have greater than 90% MTase activity, and, hopefully, more than 100%. This variant, deemed as a candidate for further study, had a Euclidean distance that was very close to the "high" activity cluster. We have very high confidence in this candidate, nevertheless, we have not completely written off the other nine new variants as new candidate possibilities because we do not have MTase data for these yet. In fact, we recommend that if it is possible, these variants should be studied in a wet lab environment to gain insights into their potential to bind nsp16 or nsp14, and this would further add to the accuracy and fine-tuning of the MDPPM. We found that the new variants created their own cluster, which was further away from the deleterious variants. It is not beyond the realm of possibilities that the variant found furthest away from the "low" MTase variants, as well as the wild type, L92Y, may be the variant that has the most potential to increase the affinity for nsp16.

The data shown in Table 2 has two variants that fall in the high category and thirteen variants that fall in the low category. This presents the potential of the data being unbalanced, e.g., the numbers of data entries in each class are not the similar. During the training and testing the random forests model displays 100% accuracy, precision and recall. As the data is not balanced, the F-measure or balanced F-score (F1 score) can be calculated as a measure of test's accuracy. In the case of this model the F1 score is 1.0 which indicated high accuracy [73]. In the case of this study each of the variants has ~30,000 structures (data points) making ~60,000 in the high category and 260,000 in the low category. When there is a large number of data points (large data set), then problems caused by unbalanced data are reduced because the minority class (H) is well represented by the large number of data points [74]. Therefore, the issue of unbalanced data does not confound the results presented here. If the accuracy of predicting the high category was closer to 87% (13/15), there would be more of a concern about the data being unbalanced.

The findings of the phi and psi angle prediction accuracy by random forest discovered that accurate angles were mostly located in the 70's and 90's portion of the peptide, suggesting that these residues are important structurally and that changes in residues here may cause a decrease in MTase activity. It should be noted that residues in the 80's portion of the peptide are mostly semi-conserved, and that they are not as conserved as those in the 70's and 90's, which may also be a contributing factor to these results. Further, we found that the chelating residues of C74, C77, and C90, located in the zinc finger, were in the top 20 phi and psi angles for accuracy predictions, and that D82, adjacent to the fourth chelating molecule H83, was in the Gini index, which calculates the data taking part in the decision making of the algorithm. If

the peptide cannot chelate to zinc it most likely loses function, therefore peptides affecting these residues or those adjacent may not be good picks for future peptide candidates.

Structurally and qualitatively, we were able to discern that the variants which had a rotamer at the H80 position had decreased MTase activity. This rotation also caused a knock-on effect which moved I81 and D82 away from the WT location. The H80 rotation is significant because even though H80 is not involved in binding nsp16, a rotation towards the C77 molecule may well affect the binding of nsp16, because C77 is directly involved in the nsp10/nsp16 interaction. And this may also be why nsp16 MTase activity is not as affected as nsp14 in the H80A variant, because it can only affect nsp16 indirectly through the C77 residue [15]. Converse to nsp16, the H80 residue is directly involved in nsp14 binding, so any movement in this case is detrimental.

The Y96"X" residue variants' most important change seemed to be the Y96 amino acid itself. Although the other variants had a slight change qualitatively in the A71 residue, which is involved in the nsp16 interaction, the molecular simulations suggest that are two possibilities for the decrease in MTase activity, either subtle changes in the entire structure add up to an overall change that affects binding, or, that Y96 is important for a different reason, possibly biochemically.

5. Conclusion

In conclusion, molecular dynamics coupled with machine learning can afford a way to predict whether new candidate peptides will have functional changes that affect methyltransferase activity. Molecular simulations show that there are structural misalignments at residues F68, H80, I81, and D94 compared to wild type that are associated with loss of MTase activity. Another structural change is the rotation near H80, where the differences that each variant creates can be visualized, which allows any subsequent variants with the same phenotype to be eliminated. However, as is the case with Y96 variants, when there are no distinct structural changes, we can rely on the specificity and accuracy of this new method. We have shown that with the nsp10 peptide we were able to predict precise residues that affect the binding of nsp16, and that these need to be considered when developing new sequences. Furthermore, this study has described a method in which to produce compelling new peptide leads computationally. For example, the H80R variant peptide is predicted to attenuate MTase activity and is a potential candidate for an anti-MTase peptide. Finally, this method shows promise in weeding out variants that may be of no use to scientists or those with limited resources who may only be able to test one or two sequences at a time. With the potential of septillions of residues in this small sequence alone, simulations coupled with machine learning can be of great use to scientists in the discovery of new drug leads.

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.

Acknowledgements

No external funding supported this work.

References

[1] Chen Y, Guo D. Molecular mechanisms of coronavirus RNA capping and methylation. Virol Sin 2016;31:3–11.
[2] Zust R, Cervantes-Barragan L, Habjan M, Maier R, Neuman BW, Ziebuhr J, Szretter KJ, Baker SC, Barchet W, Diamond MS, Siddell SG, Ludewig B, Thiel V. Ribose 2'-O-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. Nat Immunol 2011;12:137–43.
[3] Romano M, Ruggiero A, Squeglia F, Maga G, Berisio R. A structural view of SARS-CoV-2 RNA replication machinery: RNA synthesis, proofreading and final capping. Cells 2020:9.
[4] Lin S, Chen H, Chen Z, Yang F, Ye F, Zheng Y, Yang J, Lin X, Sun H, Wang L, Wen A, Dong H, Xiao Q, Deng D, Cao Y, Lu G. Crystal structure of SARS-CoV-2 nsp10 bound to nsp14-ExoN domain reveals an exoribonuclease with both structural and functional integrity. Nucleic Acids Res 2021;49:5382–92.
[5] Kozielski F, Sele C, Talibov VO, Lou J, Dong D, Wang Q, Shi X, Nyblom M, Rogstam A, Krojer T, Fisher Z, Knecht W. Identification of fragments binding to SARS-CoV-2 nsp10 reveals ligand-binding sites in conserved interfaces between nsp10 and nsp14/nsp16. RSC Chemical Biology; 2022.
[6] Bobrovs R, Kanepe I, Narvaiss N, Patetko L, Kalnins G, Sisovs M, Bula AL, Grinberga S, Boroduskis M, Ramata-Stunda A, Rostoks N, Jirgensons A, Tars K, Jaudzems K. Discovery of SARS-CoV-2 Nsp14 and Nsp16 methyltransferase inhibitors by high-throughput virtual screening. Pharmaceuticals 2021;14:1243.
[7] Bobiļeva O, Bobrovs R, Kaņepe I, Patetko L, Kalniņš G, Šišovs M, Bula AL, Gri Nberga S, Boroduškis MR, Ramata-Stunda A, Rostoks N, Jirgensons A, Ta Rs K, Jaudzems K. Potent SARS-CoV-2 mRNA cap methyltransferase inhibitors by bioisosteric replacement of methionine in SAM cosubstrate. ACS Med Chem Lett 2021;12:1102–7.
[8] Devkota K, Schapira M, Perveen S, Khalili Yazdi A, Li F, Chau I, Ghiabi P, Hajian T, Loppnau P, Bolotokova A, Satchell KJF, Wang K, Li D, Liu J, Smil D, Luo M, Jin J, Fish PV, Brown PJ, Vedadi M. Probing the SAM binding site of SARS-CoV-2 Nsp14 in vitro using SAM competitive inhibitors guides developing selective bisubstrate inhibitors. SLAS DISCOVERY: Adv Sci Drug Discov 2021;26:1200–11.
[9] Basu S, Mak T, Ulferts R, Wu M, Deegan T, Fujisawa R, Tan KW, Lim CT, Basier C, Canal B, Curran JF, Drury LS, McClure AW, Roberts EL, Weissmann F, Zeisner TU, Beale R, Cowling VH, Howell M, Labib K, Diffley JFX. Identifying SARS-CoV-2 antiviral compounds by screening for small molecule inhibitors of Nsp14 RNA cap methyltransferase. Biochem J 2021;478:2481–97.
[10] Chen Y, Tao J, Sun Y, Wu A, Su C, Gao G, Cai H, Qiu S, Wu Y, Ahola T, Guo D. Structure-function analysis of severe acute respiratory syndrome coronavirus RNA cap guanine-N7-methyltransferase. J Virol 2013;87:6296–305.
[11] Wang Y, Sun Y, Wu A, Xu S, Pan R, Zeng C, Jin X, Ge X, Shi Z, Ahola T, Chen Y, Guo D. Coronavirus nsp10/nsp16 methyltransferase can Be targeted by nsp10-derived peptide in vitro and in vivo to reduce replication and pathogenesis. J Virol 2015;89:8416–27.
[12] Ma Y, Wu L, Shaw N, Gao Y, Wang J, Sun Y, Lou Z, Yan L, Zhang R, Rao Z. Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. Proc Natl Acad Sci U S A 2015;112:9436–41.
[13] Ke M, Chen Y, Wu A, Sun Y, Su C, Wu H, Jin X, Tao J, Wang Y, Ma X, Pan JA, Guo D. Short peptides derived from the interaction domain of SARS coronavirus nonstructural protein nsp10 can suppress the 2'-O-methyltransferase activity of nsp10/nsp16 complex. Virus Res 2012;167:322–8.
[14] Hsu JC-C, Laurent-Rolle M, Pawlak JB, Wilen CB, Cresswell P. Translational shutdown and evasion of the innate immune response by SARS-CoV-2 NSP14 protein. Proc Natl Acad Sci Unit States Am 2021;118:e2101161118.
[15] Bouvet M, Lugari A, Posthuma CC, Zevenhoven JC, Bernard S, Betzi S, Imbert I, Canard B, Guillemot JC, Lecine P, Pfefferle S, Drosten C, Snijder EJ, Decroly E, Morelli X. Coronavirus Nsp10, a critical co-factor for activation of multiple replicative enzymes. J Biol Chem 2014;289:25783–96.
[16] Lugari A, Betzi S, Decroly E, Bonnaud E, Hermant A, Guillemot JC, Debarnot C, Borg JP, Bouvet M, Canard B, Morelli X, Lecine P. Molecular mapping of the RNA Cap 2'-O-methyltransferase activation interface between severe acute respiratory syndrome coronavirus nsp10 and nsp16. J Biol Chem 2010;285:33230–41.
[17] Aouadi W, Eydoux C, Coutard B, Martin B, Debart F, Vasseur JJ, Contreras JM, Morice C, Querat G, Jung ML, Canard B, Guillemot JC, Decroly E. Toward the identification of viral cap-methyltransferase inhibitors by fluorescence screening assay. Antivir Res 2017;144:330–9.
[18] Lau JL, Dunn MK. Therapeutic peptides: historical perspectives, current development trends, and future directions. Bioorg Med Chem 2018;26:2700–7.
[19] Lee AC, Harris JL, Khanna KK, Hong JH. A comprehensive review on current advances in peptide drug development and design. Int J Mol Sci 2019;20.
[20] Craik DJ, Fairlie DP, Liras S, Price D. The future of peptide-based drugs. Chem Biol Drug Des 2013;81:136–47.
[21] Ahmed A, Siman-Tov G, Hall G, Bhalla N, Narayanan A. Human antimicrobial peptides as therapeutics for viral infections. Viruses 2019:11.
[22] Lei J, Sun L, Huang S, Zhu C, Li P, He J, Mackey V, Coy DH, He Q. The antimicrobial peptides and their potential clinical applications. Am J Transl Res 2019;11:3919–31.
[23] Feng M, Fei S, Xia J, Labropoulou V, Swevers L, Sun J. Antimicrobial peptides as potential antiviral factors in insect antiviral immune response. Front Immunol 2020;11:2030.
[24] Muttenthaler M, King GF, Adams DJ, Alewood PF. Trends in peptide drug discovery. Nat Rev Drug Discov 2021;20:309–25.
[25] Guo Z. The modification of natural products for medical use. Acta Pharm Sin B 2017;7:119–36.
[26] Pelkonen O, Turpeinen M, Raunio H. In vivo-in vitro-in silico pharmacokinetic modelling in drug development: current status and future directions. Clin Pharmacokinet 2011;50:483–91.
[27] Noé F, Tkatchenko A, Müller K-R, Clementi C. Machine learning for molecular simulation. Annu Rev Phys Chem 2020;71:361–90.
[28] Agrahari AK, Krishna Priya M, Praveen Kumar M, Tayubi IA, Siva R, Prabhu Christopher B, George Priya Doss C, Zayed H. Understanding the structure-function relationship of HPRT1 missense mutations in association with Lesch-Nyhan disease

and HPRT1-related gout by in silico mutational analysis. Comput Biol Med 2019; 107:161–71.
[29] Sinha S, Wang SM. Classification of VUS and unclassified variants in BRCA1 BRCT repeats by molecular dynamics simulation. Comput Struct Biotechnol J 2020;18: 723–36.
[30] Gebhardt J, Kiesel M, Riniker S, Hansen N. Combining molecular dynamics and machine learning to predict self-solvation free energies and limiting activity coefficients. J Chem Inf Model 2020;60:5319–30.
[31] Kumar A, Purohit R. Use of long term molecular dynamics simulation in predicting cancer associated SNPs. PLoS Comput Biol 2014;10:e1003318.
[32] Wang DD, Ou-Yang L, Xie H, Zhu M, Yan H. Predicting the impacts of mutations on protein-ligand binding affinity based on molecular dynamics simulations and machine learning methods. Comput Struct Biotechnol J 2020;18:439–54.
[33] Jamal S, Grover A, Grover S. Machine learning from molecular dynamics trajectories to predict caspase-8 inhibitors against alzheimer's disease. Front Pharmacol 2019;10:780.
[34] McCoy MD, Hamre 3rd J, Klimov DK, Jafri MS. Predicting genetic variation severity using machine learning to interpret molecular simulations. Biophys J; 2020.
[35] Hamre III JR, Klimov DK, McCoy MD, Jafri MS. Machine learning-based prediction of drug and ligand binding in BCL-2 variants through molecular dynamics. Comput Biol Med 2022;140:105060.
[36] Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem 2005;26:1781–802.
[37] Phillips JC, Hardy DJ, Maia JDC, Stone JE, Ribeiro JV, Bernardi RC, Buch R, Fiorin G, Henin J, Jiang W, McGreevy R, Melo MCR, Radak BK, Skeel RD, Singharoy A, Wang Y, Roux B, Aksimentiev A, Luthey-Schulten Z, Kale LV, Schulten K, Chipot C, Tajkhorshid E. Scalable molecular dynamics on CPU and GPU architectures with NAMD. J Chem Phys 2020;153:044130.
[38] Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmuller H, MacKerell Jr AD. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods 2017;14:71–3.
[39] Kobayashi H, Picard LP, Schonegge AM, Bouvier M. Bioluminescence resonance energy transfer-based imaging of protein-protein interactions in living cells. Nat Protoc 2019;14:1084–107.
[40] Wren LM, Jimenez-Jaimez J, Al-Ghamdi S, Al-Aama JY, Bdeir A, Al-Hassnan ZN, Kuan JL, Foo RY, Potet F, Johnson CN, Aziz MC, Carvill GL, Kaski JP, Crotti L, Perin F, Monserrat L, Burridge PW, Schwartz PJ, Chazin WJ, Bhuiyan ZA, George Jr AL. Genetic mosaicism in calmodulinopathy. Circ Genom Precis Med 2019;12:375–85.
[41] Huan Y, Kong Q, Mou H, Yi H. Antimicrobial peptides: classification, design, application and research progress in multiple fields. Front Microbiol 2020;11.
[42] Müller AT, Hiss JA, Schneider G. Recurrent neural network model for constructive peptide design. J Chem Inf Model 2018;58:472–9.
[43] Nagarajan D, Nagarajan T, Roy N, Kulkarni O, Ravichandran S, Mishra M, Chakravortty D, Chandra N. Computational antimicrobial peptide design and evaluation against multidrug-resistant clinical isolates of bacteria. J Biol Chem 2018;293:3492–509.
[44] Tucs A, Tran DP, Yumoto A, Ito Y, Uzawa T, Tsuda K. Generating ampicillin-level antimicrobial peptides with activity-aware generative adversarial networks. ACS Omega 2020;5:22847–51.
[45] Dean SN, Alvarez JAE, Zabetakis D, Walper SA, Malanoski AP. PepVAE: variational autoencoder framework for antimicrobial peptide generation and activity prediction. Front Microbiol 2021;12. 725727-725727.
[46] Mulligan VK. The emerging role of computational design in peptide macrocycle drug discovery. Expet Opin Drug Discov 2020;15:833–52.
[47] Playe B, Azencott CA, Stoven V. Efficient multi-task chemogenomics for drug specificity prediction. PLoS One 2018;13:e0204999.
[48] Yan Z, Wang J. Specificity quantification of biomolecular recognition and its implication for drug discovery. Sci Rep 2012;2:309.
[49] Boothroyd S, Kerridge A, Broo A, Buttar D, Anwar J. Solubility prediction from first principles: a density of states approach. Phys Chem Chem Phys : Phys Chem Chem Phys 2018;20:20981–7.
[50] Bergström CAS, Larsson P. Computational prediction of drug solubility in water-based systems: qualitative and quantitative approaches used in the current drug discovery and development setting. Int J Pharm 2018;540:185–93.
[51] Egan WJ, Lauri G. Prediction of intestinal permeability. Adv Drug Deliv Rev 2002; 54:273–89.
[52] Dickson CJ, Hornak V, Pearlstein RA, Duca JS. Structure-kinetic relationships of passive membrane permeation from multiscale modeling. J Am Chem Soc 2017; 139:442–52.
[53] Zhang L, Zhang H, Ai H, Hu H, Li S, Zhao J, Liu H. Applications of machine learning methods in drug toxicity prediction. Curr Top Med Chem 2018;18:987–97.
[54] Li X, Chen L, Cheng F, Wu Z, Bian H, Xu C, Li W, Liu G, Shen X, Tang Y. In silico prediction of chemical acute oral toxicity using multi-classification methods. J Chem Inf Model 2014;54:1061–9.
[55] Agrawal P, Singh H, Srivastava HK, Singh S, Kishore G, Raghava GPS. Benchmarking of different molecular docking methods for protein-peptide docking. BMC Bioinf 2019;19:426.
[56] Can H, Koseoglu AE, Erkunt Alak S, Guvendi M, Doskaya M, Karakavuk M, Guruz AY, Un C. In silico discovery of antigenic proteins and epitopes of SARS-CoV-2 for the development of a vaccine or a diagnostic approach for COVID-19. Sci Rep 2020;10:22387.
[57] Sk MF, Jonniya NA, Roy R, Poddar S, Kar P. Computational investigation of structural dynamics of SARS-CoV-2 methyltransferase-stimulatory factor heterodimer nsp16/nsp10 bound to the cofactor SAM. Front Mol Biosci 2020;7: 590165.
[58] Wang J. Fast identification of possible drug treatment of coronavirus disease-19 (COVID-19) through computational drug repurposing study. J Chem Inf Model 2020;60:3277–86.
[59] Ling R, Dai Y, Huang B, Huang W, Yu J, Lu X, Jiang Y. In silico design of antiviral peptides targeting the spike protein of SARS-CoV-2. Peptides 2020;130:170328.
[60] Sitthiyotha T, Chunsrivirot S. Computational design of 25-mer peptide binders of SARS-CoV-2. J Phys Chem B 2020;124:10930–42.
[61] Wang Z, Sun H, Yao X, Li D, Xu L, Li Y, Tian S, Hou T. Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power. Phys Chem Chem Phys : Phys Chem Chem Phys 2016;18:12964–75.
[62] Este JA, Cihlar T. Current status and challenges of antiretroviral research and therapy. Antivir Res 2010;85:25–33.
[63] Otvos Jr L, Wade JD. Current challenges in peptide-based drug discovery. Front Chem 2014;2:62.
[64] Powell MF, Grey H, Gaeta F, Sette A, Colon S. Peptide stability in drug development: a comparison of peptide reactivity in different biological media. J Pharmacol Sci 1992;81:731–5.
[65] Dang A, Vallish BN, Dang S. Hydroxychloroquine and Remdesivir in COVID-19: a critical analysis of recent events. Indian J Med Ethics 2020;V:202–7.
[66] Million M, Lagier JC, Gautret P, Colson P, Fournier PE, Amrane S, Hocquart M, Mailhe M, Esteves-Vieira V, Doudier B, Aubry C, Correard F, Giraud-Gatineau A, Roussel Y, Berenger C, Cassir N, Seng P, Zandotti C, Dhiver C, Ravaux I, Tomei C, Eldin C, Tissot-Dupont H, Honore S, Stein A, Jacquier A, Deharo JC, Chabriere E, Levasseur A, Fenollar F, Rolain JM, Obadia Y, Brouqui P, Drancourt M, La Scola B, Parola P, Raoult D. Early treatment of COVID-19 patients with hydroxychloroquine and azithromycin: a retrospective analysis of 1061 cases in Marseille, France. Trav Med Infect Dis 2020;35:101738.
[67] Grein J, Ohmagari N, Shin D, Diaz G, Asperges E, Castagna A, Feldt T, Green G, Green ML, Lescure FX, Nicastri E, Oda R, Yo K, Quiros-Roldan E, Studemeister A, Redinski J, Ahmed S, Bernett J, Chelliah D, Chen D, Chihara S, Cohen SH, Cunningham J, D'Arminio Monforte A, Ismail S, Kato H, Lapadula G, L'Her E, Maeno T, Majumder S, Massari M, Mora-Rillo M, Mutoh Y, Nguyen D, Verweij E, Zoufaly A, Osinusi AO, DeZure A, Zhao Y, Zhong L, Chokkalingam A, Elboudwarej E, Telep L, Timbs L, Henne I, Sellers S, Cao H, Tan SK, Winterbourne L, Desai P, Mera R, Gaggar A, Myers RP, Brainard DM, Childs R, Flanigan T. Compassionate use of remdesivir for patients with severe covid-19. N Engl J Med 2020;382:2327–36.
[68] Krafcikova P, Silhan J, Nencka R, Boura E. Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin. Nat Commun 2020;11:3717.
[69] Ray D, Shah A, Tilgner M, Guo Y, Zhao Y, Dong H, Deas TS, Zhou Y, Li H, Shi PY. West Nile virus 5'-cap structure is formed by sequential guanine N-7 and ribose 2'-O methylations by nonstructural protein 5. J Virol 2006;80:8362–70.
[70] Zhou Y, Ray D, Zhao Y, Dong H, Ren S, Li Z, Guo Y, Bernard KA, Shi PY, Li H. Structure and function of flavivirus NS5 methyltransferase. J Virol 2007;81: 3891–903.
[71] Decroly E, Imbert I, Coutard B, Bouvet M, Selisko B, Alvarez K, Gorbalenya AE, Snijder EJ, Canard B. Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (Nucleoside-2′O)-Methyltransferase activity. J Virol 2008;82: 8071–84.
[72] Dutta M, Iype E. Peptide inhibitors against SARS-CoV-2 2′-O-methyltransferase involved in RNA capping: a computational approach. Biochem Biophys Rep 2021; 27:101069.
[73] Hripcsak G, Rothschild AS. Agreement, the f-measure, and reliability in information retrieval. J Am Med Inf Assoc 2005;12:296–8.
[74] Batista GEAPA, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explor 2004;6:20–9.