ORIGINAL ARTICLE

A multimodal deep learning-based drug repurposing approach for treatment of COVID-19

Seyed Aghil Hooshmand[1,2] · Mohadeseh Zarei Ghobadi[2] · Seyyed Emad Hooshmand[3] ·
Sadegh Azimzadeh Jamalkandi[4] · Seyed Mehdi Alavi[5] · Ali Masoudi-Nejad[1,2]

Abstract

Recently, various computational methods have been proposed to find new therapeutic applications of the existing drugs. The Multimodal Restricted Boltzmann Machine approach (MM-RBM), which has the capability to connect the information about the multiple modalities, can be applied to the problem of drug repurposing. The present study utilized MM-RBM to combine two types of data, including the chemical structures data of small molecules and differentially expressed genes as well as small molecules perturbations. In the proposed method, two separate RBMs were applied to find out the features and the specific probability distribution of each datum (modality). Besides, RBM was used to integrate the discovered features, resulting in the identification of the probability distribution of the combined data. The results demonstrated the significance of the clusters acquired by our model. These clusters were used to discover the medicines which were remarkably similar to the proposed medications to treat COVID-19. Moreover, the chemical structures of some small molecules as well as dys-regulated genes' effect led us to suggest using these molecules to treat COVID-19. The results also showed that the proposed method might prove useful in detecting the highly promising remedies for COVID-19 with minimum side effects. All the source codes are accessible using https://github.com/LBBSoft/Multimodal-Drug-Repurposing.git

✉ Ali Masoudi-Nejad
amasoudin@ut.ac.ir

1 Laboratory of Systems Biology and Bioinformatics (LBB), Department of Bioinformatics, University of Tehran, Kish International Campus, Kish Island, Iran

2 Laboratory of Systems Biology and Bioinformatics (LBB), Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

3 Department of Medical Nanotechnology, Faculty of Advanced Technologies in Medicine, Iran University of Medical Sciences, Tehran, Iran

4 Chemical Injuries Research Center, Systems Biology and Poisonings Institute, Tehran, Iran

5 Department of Plant Biotechnology, National Institute of Genetic Engineering and Biotechnology, Tehran, Iran

Graphic abstract
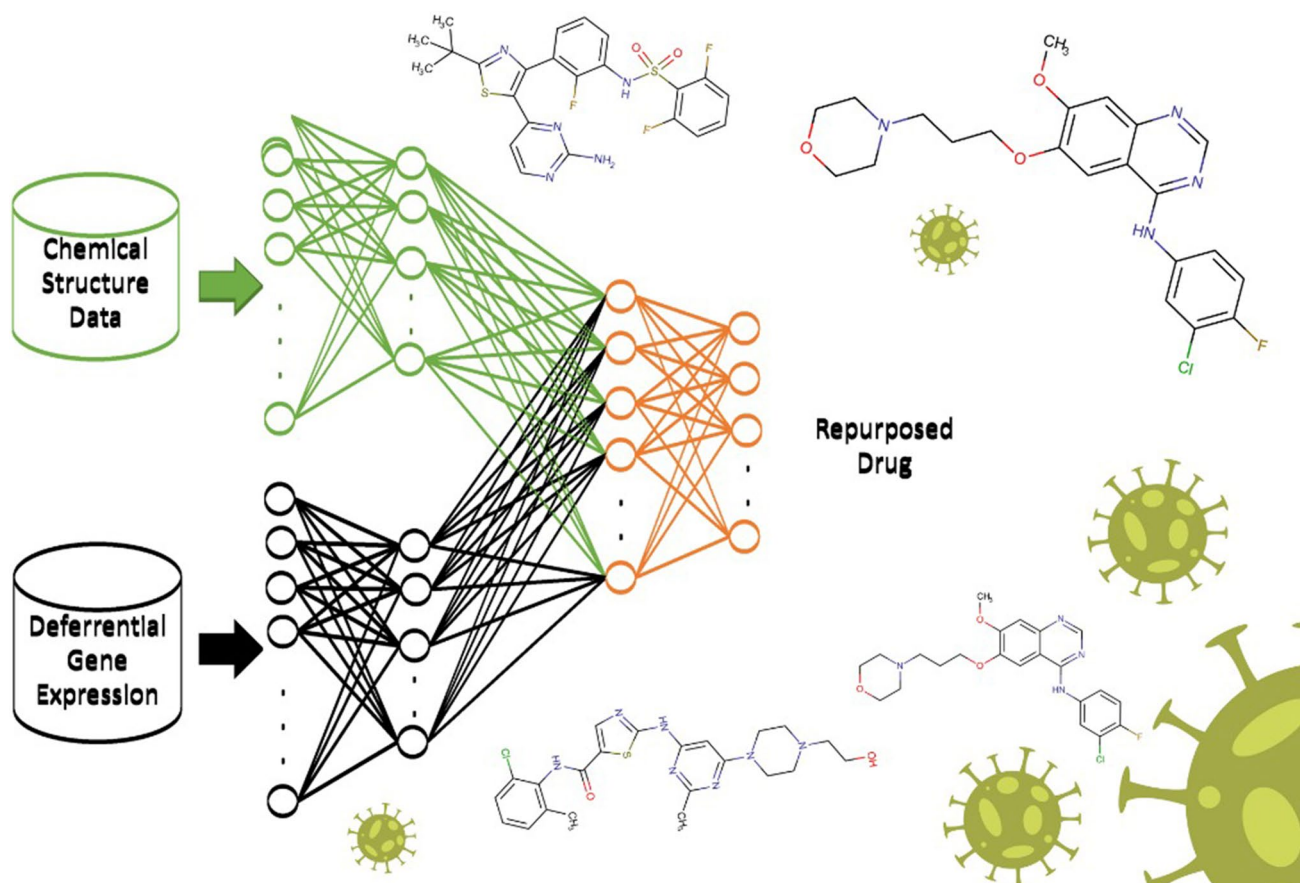
Introduction

COVID-19 pandemic, which became a trigger for a series of serious challenges, was brought about by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in December 2019. Until now, various FDA-approved and clinical trialed drugs have been suggested for treating the disease. Five antiviral drugs, including chloroquine phosphate, interferon α (IFN-α), lopinavir/ritonavir, arbidol, and ribavirin from among prime candidate drugs have been introduced for treating COVID-19.

Drug repurposing is regarded as a way to discover new applications of the existing drugs in treating various diseases [1]. This technique may speed up the process of finding the therapeutic compounds for the newly emerged illnesses [2–6]. Signature matching, which is based on detecting the unique signatures of drugs, is one of the computational approaches to be used. Chemical structures, transcriptomic (RNA), and proteomic or metabolomic data, and adverse event profiles are some samples of drug signatures [1]. The

similarity between the drug-disease and drug-drug predictions can be identified by matching transcriptomic signatures [7, 8]. To discover the drug-disease similarity to drug-drug, the differentially expressed genes (DEGs) are first measured before and after using drugs for a cell line or a tissue and, then, compared with DEGs which correspond with healthy samples. Moreover, considering the hypothesis of the similar therapeutic mechanisms of drugs, various signatures can be taken into account when determining the drug-drug's similarity to drug-disease [9].

Deep learning, which utilizes artificial neural networks (ANN) concepts, is a branch of machine learning techniques. One of the main benefits of deep learning is its ability to automatically detect relationships between features and form a multilayer stack of neural networks with nonlinear input–output mappings [10].

Compared with the other machine learning methods, multimodal learning, which connects the existing information about multiple modalities, has many advantages such as representing features, supporting different states of fusion,

learning the architecture of fusion, and expanding based on the number of modalities and data sizes [11]. In the multimodal fusion setting, all of the modalities data are accessible in the whole phases [12].

Considering the spread of SARS-CoV-2 at lightning speed, there is an urgent need to find promising drugs to inhibit the virus from spreading and control COVID-19. As part of our ongoing program associated with developing computational methods as well as drug repurposing [3, 6, 13], we aimed to introduce a multimodal RBM approach, named MM-RBM, to find the drugs that are very similar to the four above-mentioned drugs tested on COVID-19. To this end, we used the data on the chemical structures of medicines as well as DEGs identified after using drugs.

Materials and methods

Preparation of datasets

Two central databases of Harmonizome and LINCS were explored to extract appropriate datasets. Harmonizome gathers the knowledge and information about genes and proteins (https://amp.pharm.mssm.edu/Harmonizome) [14]. It contains 114 datasets on different subjects; one of them comprises the differentially expressed genes (DEGs) along with small molecules perturbations retrieved from the LINCS (Literacy Information and Communication System) database [15]. All of the small molecules, which were injected into **cells** in different doses and time points, were downloaded (4 million records). After excluding the unreadable data and N/A values, first, a matrix containing small molecules (applied in various doses and time points) was constructed in rows and then genes were formed in columns. In the constructed matrix, "1", "−1", and "0" were assigned to the up-regulated genes, down-regulated genes, and unchanged genes, respectively. To create a universal ID for all of the data, the small molecules' IDs (BIRD-ID) were converted into the PubChem Compound ID (CID). For this purpose, all the 40,000 small molecules, deposited in the LINCS project, were downloaded. Then, LINCS-ID, PubChem-ID, and BIRD-ID were specified for each fragment. Since there existed no PubChem-IDs for 77 BIRD-IDs, they were removed in the further data analysis. After testing various dimension reduction methods, variance calculations were

selected and the columns (DEGs), i.e., the third quantile of variances, with higher values were maintained.

Moreover, all the small molecules' SMILE codes which represent the chemical features were downloaded from PubChem. MACC fingerprint, which is a static binary **of** each molecule, was also determined using the PADEL package. MACC fingerprint may be more useful than the other static fingerprints such as PubChem, KRFP, and some dynamic fingerprints like ECFP [13]. The columns with zero variance were removed too.

Table 1 shows the number of the components of drug-DEGs and drug–chemical features data matrices before and after preprocessing. First, to acquire an expression data matrix for 29,074 small molecules, the binary data between the small molecules and the genes were downloaded from the Harmonizome database. Next, a matrix with 8,000 genes was obtained. Considering the large number of features, the upper third quantile of variances was chosen as features. To acquire the chemical features of these molecules, first, their SMILE was downloaded, and, then, 166 MACC properties were calculated for each molecule. Consequently, a matrix containing 166 columns was created. However, the columns whose variances were zero were removed. The end result of the above-mentioned process was a matrix with 156 columns, which was regarded as the final matrix. The elements of both matrices were assigned 0 or 1.

Restricted Boltzmann machine

The restricted Boltzmann machine (RBM) is a popular unsupervised stochastic neural network. When a neuron is activated, it will show random behavior. The binary RBM contains both the visible (v_i) and hidden (h_j) layers that are fully connected via pair-wise potentials. However, the limitation of no within-layer connection is applied to both the visible and hidden layers [16]. The distribution of the hidden units h, given the visible units v, is as follow:

$$p(h \vee v; \theta) = \prod_j p(h_j \vee v)$$

and the distribution of the hidden units v, given the visible units h, is as follow:

Table 1 Characteristics of drug-DEGs and drug–chemical features data matrixes

Number of DEGs (columns)	Number of molecules (rows)	Number of DEGs (columns)	Number of molecules (rows)	Data type
After preprocessing		Before preprocessing		
2086	29,074	8351	29,074	Drug-DEGs data
16	29,074	166	29,074	Drug–chemical data

$$p(v \vee h; \theta) = \prod_i p(v_i \vee h)$$

The conditional distributions of the binary RBM are also specified as follow:

$$p(h_j = 1|v) = \sigma\left(\sum_i W_{ij} v_i + a_j\right)$$

$$p(v_i = 1|h) = \sigma\left(\sum_j W_{ij} h_j + b_i\right)$$

in which a_j and b_i are the biases of visible and hidden units, respectively, W_{ij} is the symmetric interaction term between the two units, and $\theta = \{w, b, a\}$ is the parameter of the model. Here, the contrastive divergence (CD) algorithm was employed to update θ during the training process. To speed up the learning process of RBM and elevate the accuracy, the number of data reconstruction iterations (k) was gradually increased using the CD algorithm (CD_K).

## Multimodal deep feature fusion

Gathering information through merging various modalities yields better results in the drug-disease and drug-target predictions [17] as well as drug repurposing [18–20]. In the present study, we employed our proposed systematic method in a probabilistic manner to elicit multimodal feature representations. In this method, to overcome the difficulties of finding correlations between different modalities which comprise various statistical properties, the layers of hidden units were placed between the modalities. Moreover,

a multimodal restricted Boltzmann machine (MM-RBM)-based framework was applied to acquire both the intra- and cross-modality relationships and identify drug clusters through information obtained from both the chemical structures and gene expressions modalities. The fusion of different modalities may provide us with supplementary information and increase the accuracy of our results [21]. After the CD learning algorithm was utilized to discover the intra- and cross-modality features of multi-platform data, the final states of the joint representation of latent features were applied to classify drugs based on the cross-platform input data. For this purpose, an integrative clustering of the multi-platform cancer data was provided [22]. Generally, this approach was a success in capturing the posterior distribution of the expository factors of the input data. Figure 1 demonstrates a multimodal deep network in which one path indicates the statistical traits of chemical features, and the other one shows those of DEGs. The top common hidden layer determines the common properties of the modalities. This method is different from the previously proposed multimodal approaches, in which first the features are extracted individually from various modalities, and then, kernel machines are employed to combine them [21, 23].

Learning method

We employed a multimodal RBM approach (MM-RBM) from among a multitude of available clustering methods [21, 22]. Our proposed MM-RBM includes two stages: (1) using RBM to encode the hidden features determined by each input modality and fusing cross-platform modalities using a joint representation of hidden variables and (2) identifying the common features. For the purpose of applying the proposed MM-RBM,

Fig. 1 Multimodal restricted Boltzmann machine
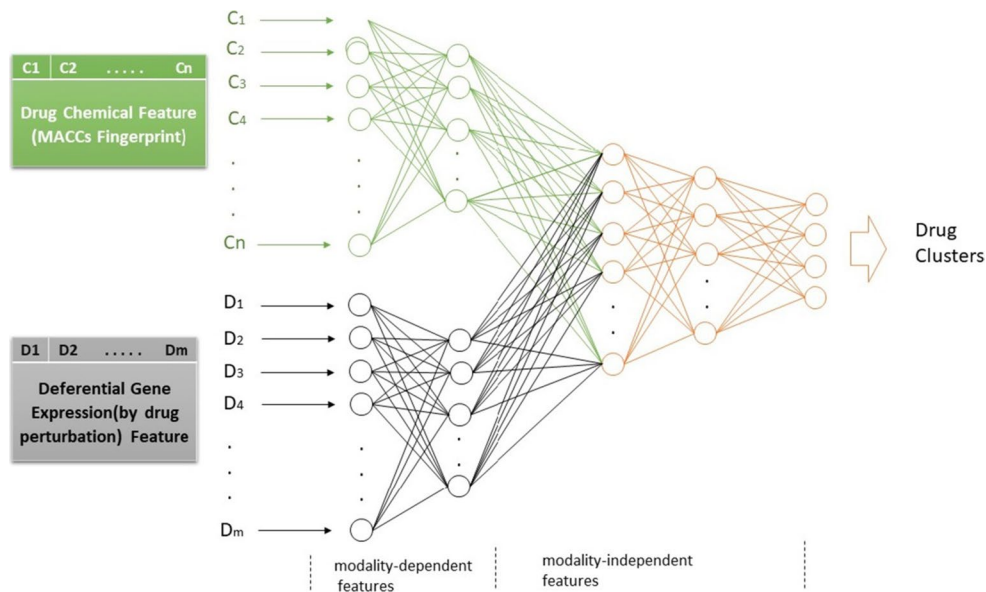
we first initialized the parameters W, b, c employing a random generator and then trained a stack of RBM layer by layer through applying the CDK algorithm to each unlabeled data (20,974 for the chemical features and DEGs data), separately. We divided the epochs into 10 sectors and elevated k one by one in each sector division. This pre-training approach was applied to both the chemical features and DEGs data twice. Next, we applied one RBM layer with 154 binary features and two stacked RBM layers with 2086 binary features as inputs in the chemical features and DEGs, respectively. The resultant values of chemical features and DEGs reconstructed by the CD algorithm were in agreement with the experimental input data, with the average correlations of 0.80 and 0.68 for the chemical features and DEGs dimensions, respectively. Since there were high correlations between the reconstructed and input data, the hidden layers in our proposed MM-RBM could display the innate features of various input data modalities.

At the end, we joined two output layers from two RBM models to fuse cross-platform modalities and achieve the shared features obtained from different input data. The CDK algorithm was used to train the stacked layers of RBM. The binary MACC fingerprints which employ the digits 0 or 1 and DEGs obtained from the Harmonizome database are the input nodes of the two single-modal RBM layers.

We conducted a grid search for hyper-parameter optimization. We also carried out research to discover an optimal number of layers as well as a number of hidden units which utilize a nested cross-validation framework. To conduct a search for hyper-parameter optimization, the number of layers should change from 1 to 5, and the number of hidden units per layer must decrease while the depths of the network are 1/2 and 1/10 in the previous layer. {154,78}, {2086,1043,104}, and {182,91} are the best combinations of parameters which we managed to obtain for chemical features data, DEGs data, and shared features, respectively. We selected RELU as a nonlinear function of hidden neurons. The best learning rate and l2 regulizers of chemical features and DEGs data were 0.01 and 0.001, respectively. Figure 2 demonstrates the reconstruction error in the RBM layers.

Implementation

The proposed method was implemented in Python 3.6. TensorFlow [24], and the open-source software library, named Keras, was also used [25]. The source codes were executed on a PC with a Corei7 CPU having eight cores, 32 GB RAM, and a NVIDIA 1070i GPU with 8 GB RAM.

Results and discussion

Clustering of data

The output of the model is categorized into 12 clusters. The number of data records and unique drugs in each cluster is shown in Table 2.

In the Harmonizome database, the gene expressions are related to different drug dosages corresponding with different time points. Hence, the number of data samples is more than the number of drugs.

Model evaluation and identification of clusters

Since the last layer consists of 4 nodes, a maximum of 16 clusters could be identified. A total of 12 clusters was obtained by applying our proposed model. To further assess the clusters, the heatmap plots representing the difference between the values 1 and 0 were depicted for all the features in each cluster. Although the separation power of clusters was found to be low by applying the chemical features, DEGs' properties helped facilitate the satisfactory separation of clusters (Fig. 3a, b). Besides, instead of applying the properties individually, applying them in groups of two or a combination of two or more properties led to the clusters' much more satisfactory separation (Fig. 3c).

We also conducted a variance analysis based on the principal component analysis (PCA). Based on the significance level ($p < 0.05$), a total of 10 and 105 significant chemical structures and DEGs features were selected, respectively. These features are described in detail in the Supplementary file 1. The PCA of the mean of MACC and DEGs indicated that all the 12 clusters display a significantly different pattern in the chemical structures and DEGs dimensions. The second principal component in cluster #2 in particular exhibited a considerable variance in the MACC data, whereas cluster #5 showed the smallest value. While the first principal component in cluster #1 indicated the largest variance in the DEG data, cluster #8 demonstrated the smallest value (Fig. 4).

To make an assessment of clusters, we used the DrugPath database [26] which contains the drug-induced pathway information obtained by changing the expression levels of genes after using drugs. The total number of unique genes which were dysregulated in each cluster is shown in Table 3.

As shown in Fig. 3c, all the 12 clusters can be categorized into 5 clusters in a manner that clusters #1, #2, #3, and #4 are put into cluster A, clusters #5, #6, and #7 into cluster B, clusters #8 and #9 into cluster C, clusters
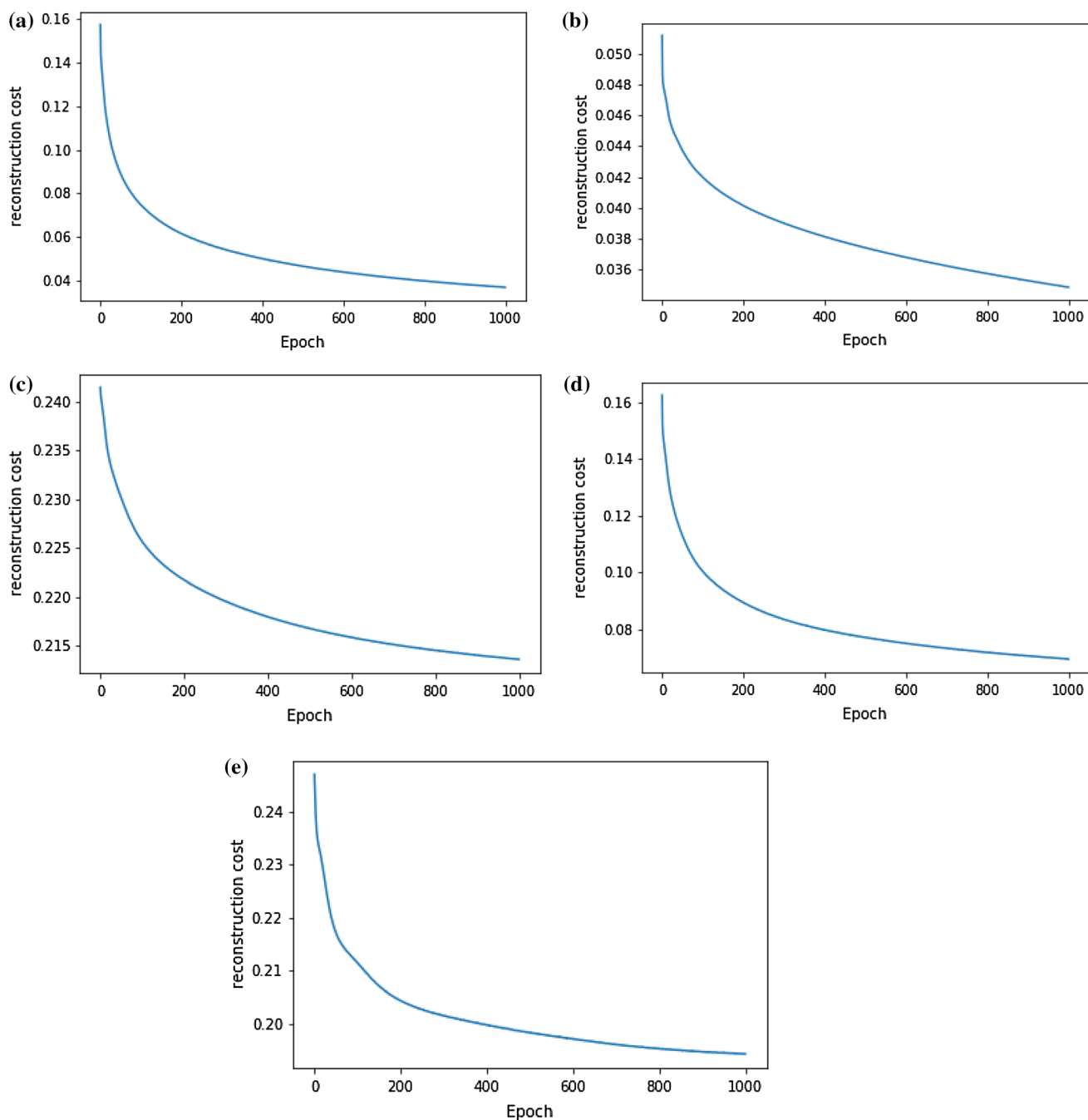
Fig. 2 The Reconstruction error in all the RBM layers. **a** The MACC hidden layer 154 × 78. **b** The Harmonizome hidden layer {2086 × 1043}. **c** The Harmonizome hidden layer {1043 × 104}. **d** The Merged MACC and Harmonizome hidden layers {182 × 91}. **e** The Merged MACC and Harmonizome hidden layers {91 × 4}

#10 and #11 into cluster D, and cluster #12 into cluster E. The distribution diagram of genes that have been dysregulated in various biological pathways through applying different drugs is shown in Fig. 5a. The Venn diagram also shows that all the new clusters except cluster B contain unique DEGs (Fig. 5b). The complete list of the specific

and common DEGs in superclusters is listed in the Supplementary file 2.

To evaluate and examine the relationship between the clusters and existing drug classes, we selected the class labels of 793 drugs of our data from the Therapeutic Use hierarchy of the Mesh Database. Sixteen classes were

Table 2 The number of different clusters obtained by applying the proposed method

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
#Record (#Drug)	(405) 537	(116)124	(1812) 4865	(1097) 2094
	Cluster 5	Cluster 6	Cluster 7	Cluster 8
#Record (#Drug)	(1281)2660	(639) 954	(1451) 3150	(399) 515
	Cluster 9	Cluster 10	Cluster 11	Cluster 12
#Record (#Drug)	(2121) 6630	(1538)3695	(1288) 2728	(744)1122

identified. Except for some classes of drugs, which belong to only a number of clusters, other classes such as anti-neoplastic agents, central nervous system, cardiovascular agents or anti-inflammatory agents emerged in almost all the clusters by the same percentage. For example, cluster #11 contains only the class of urological drugs. Smoking and antiemetic agents' classes belong to clusters #1 and #8, respectively. Although the dermatology class is distributed in all the clusters, its drugs can be found in cluster #2. Also, about a half of the hematologic agents has been included in clusters #2 and #6. The full picture of this study is given in the supplementary file 5.
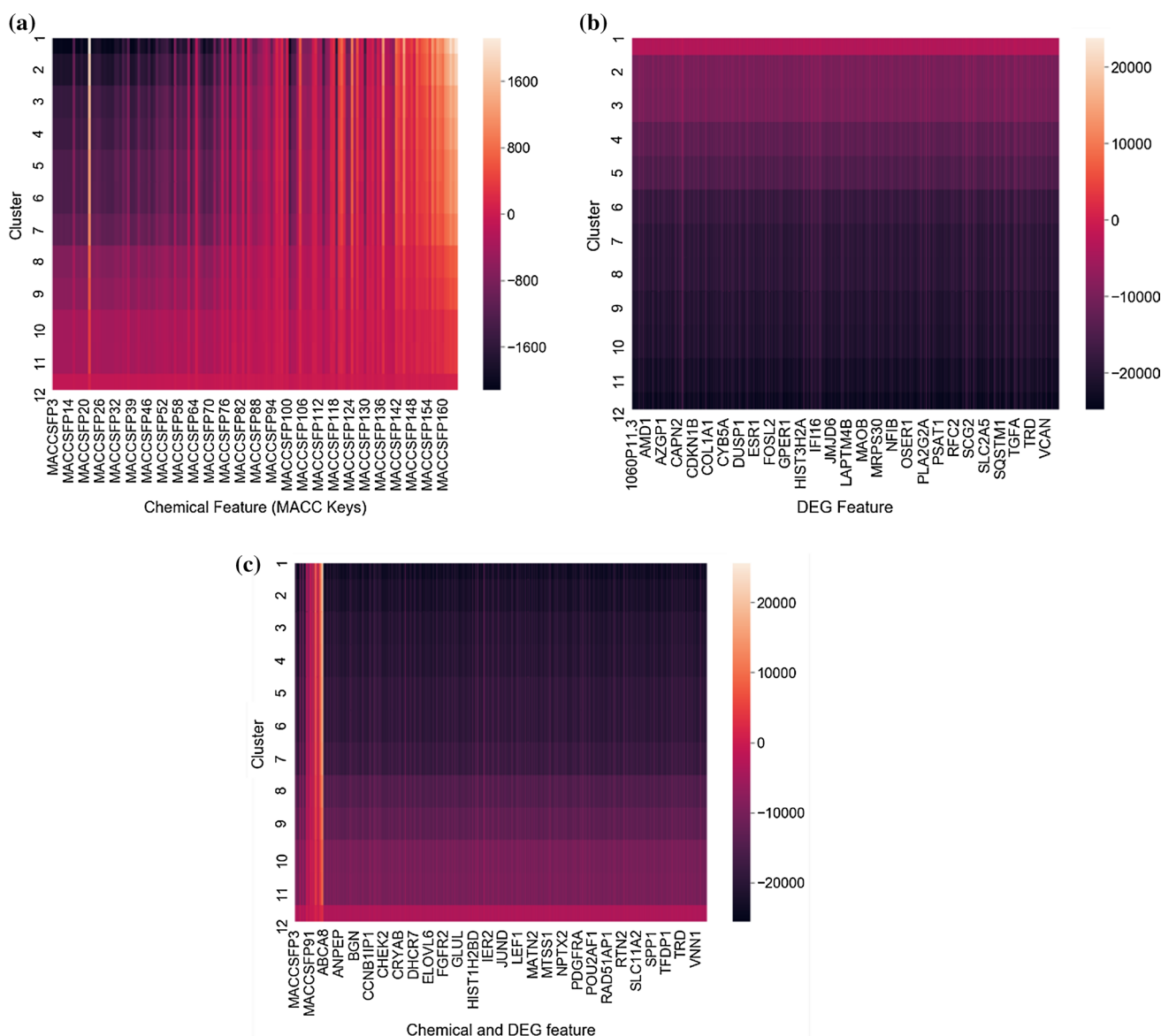
Fig. 3 The heatmap plots of clusters based on (a) the chemical structure features, b) DEGs features, and c) both the chemical structures and DEGs features

(a)

(b)
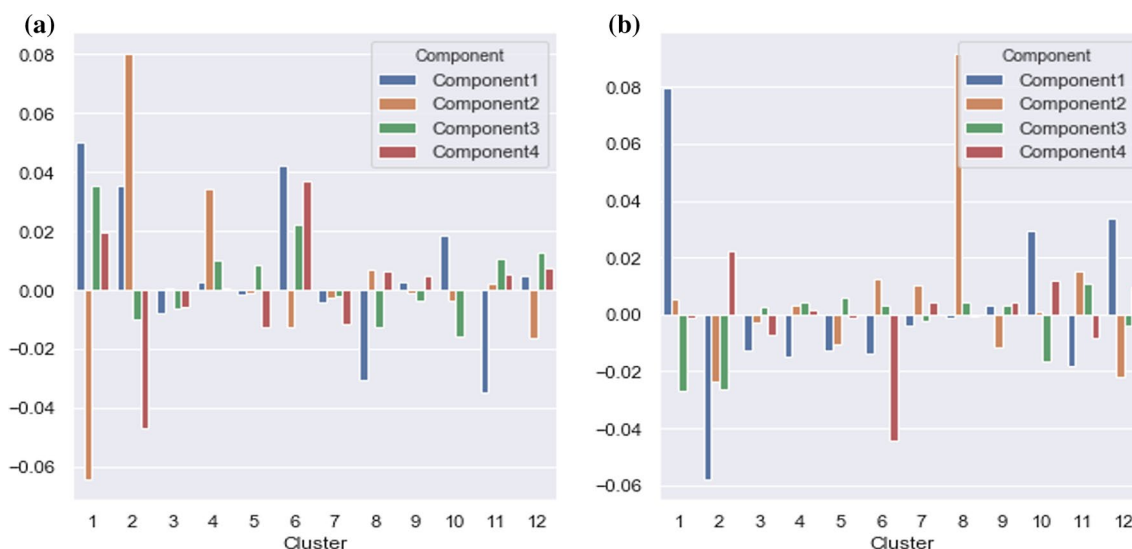
Fig. 4 The PCA plots for **a** the MACC and **b** DEGs data

Table 3 The total number of specific dysregulated genes in each cluster

Cluster	# Specific dysregulated gene	Example genes
3	21	5641+, 4041−
4	8	5322+, 10,652−
5	5	10,452+, 567−
6	3	2956+, 1270+
7	6	9846+, 2568−
9	56	661+,779−
10	17	5729+, 7100−
11	9	8986,50,617−
12	3	971+, 57,804−

Comparing multimodal and unimodal

As it has been emphasized in various works regarding drug repurposing, data integration yields better results [27]. Although using exclusively the chemical data results in the classification of strikingly similar drugs in one category, one cannot necessarily replace the other In other words, while two drugs can have completely similar chemical structures, there may be great differences between them in terms of their effectualness in treating the disease. Pharbital, for example, can produce an entirely different effect with a slight change in its chemical structure [28]. Therefore, although compared to the integrated data, the internal evaluation criteria yield better results on the chemical data types, it cannot be sufficient. Further, the addition of the expression data types can produce more accurate and sensible biological results. However, we tested our method on exclusively chemical data

types, with the result that it was established in 15 clusters. To make a comparison between multimodal and unimodal data, we used the Tanimoto criterion as an internal criterion to find out the average of both cohesion within and separation of clusters. By calculations, cohesion within single and combination data equaled 0.53 and 0.63, respectively.

Moreover, the calculated separation values of single and combination data were 0.63 and 0.65, respectively (lower cohesion and higher separation are desirable). The resultant values indicated that although the clusters of exclusively chemical data appear to have more internal cohesion compared to the other ones, there are no marked differences in clusters' behavior. Based on the outcomes of our method, due to remarkable similarity between DEGs, which are located in the same cluster, chemical data cohesion is low, and one may replace the other.

Drug repurposing for COVID-19

The COVID-19 epidemic outbreak provided the impetus for many studies which have been carried out to find the most effective treatment. To this end, scientists and medical specialists have proposed and tested a number of drugs which are usually used in the treatment of other diseases [29–33]. In the present study, first, we selected hydroxychloroquine, lopinavir, disulfiram, and nitazoxanid from among prime drugs because the Harmonizome database contains the genes expressions information as well as DEGs which were identified in the course of treatment with the above-mentioned drugs. Then, we searched for similar drugs. As shown in Fig. 6, hydroxychloroquine, lopinavir, disulfiram, and nitazoxanide fit into cluster #5, clusters #3 and #10, clusters #3 and #10, and clusters
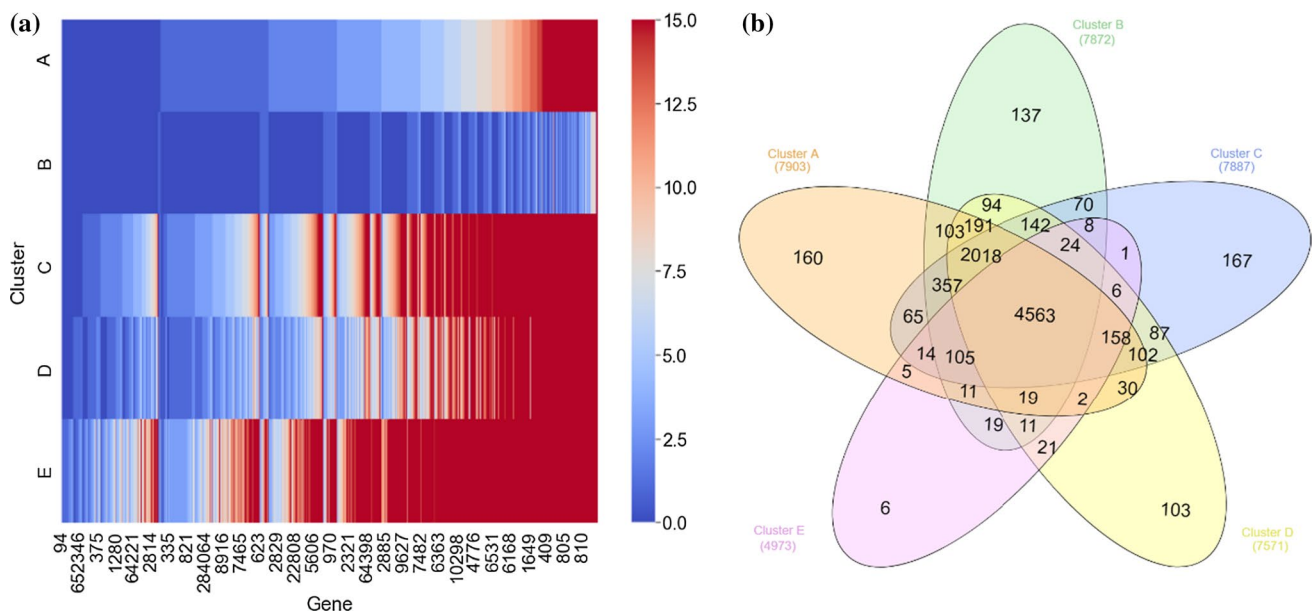
Fig. 5 **a** The heatmap of DEGs in each cluster; **b** Venn diagram contains the common and unique number of DEGs in each cluster

Fig. 6 Drugs that are similar to four potential drugs used for COVID-19
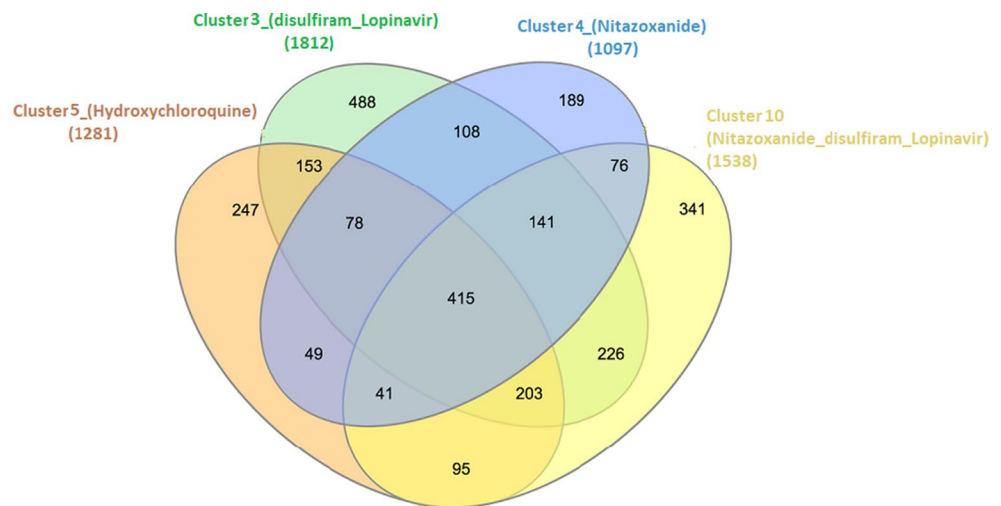
#4 and #10, respectively. Each of these clusters contains drugs that are classified in terms of the existing similarity between the chemical structure features and DEGs. A total of 415 drugs was identified; these drugs produce effects resembling those generated by the four above-mentioned drugs. A list of drugs in each cluster is presented in the Supplementary file 3.

After treating the diseases with these drugs, we identified the similarities between every two drugs in each cluster in terms of the chemical features and the dysregulated genes. As shown below, we utilized the Tanimoto coefficient to determine the chemical features similarities:

$$T(a,b) = \frac{N_a}{N_a + N_b - N_c}$$

where n represents the number of chemical features in each drug (a, b) and (c) demonstrates the intersection set.

Moreover, to determine the DEGs' similarities in every pair of drugs, we used a similarity measure based on the dysregulated genes in which each gene's score was computed:

$$sim(a_i, b_i) = \begin{cases} 0 & \text{if } a_i = b_i = 0 \\ +2 & \text{if } a_i = b_i = +1 \text{ or } a_i = b_i = -1 \\ -1 & \text{if } (a_i = 0 \text{ and } |b_i| = 1) \text{ or } (|a_i| = 1 \text{ and } b_i = 0 \\ -2 & a_i * b_i = -1 \end{cases}$$

At the end, the total number of genes in every pair of drugs was calculated using the following equation:

$$SIM(a, b) = \sum_i sim(a_i, b_i)$$

where a and b represent the drugs, a_i is the i' feature in the drug (a), and n is the number of genes in each drug.

Through applying the similarity measures, we identified the drugs that closely resembled the representative of each cluster. With regard to the chemical structures and drugs' effects on the gene expressions, we picked the drugs which bore a remarkable similarity to each drug in each cluster. Then, we examined the Comparative Toxicogenomics Database (CTD) [34] to further evaluate the accuracy of the selected drugs. Figure 7 shows that a majority of drugs are linked to COVID-19 by at least one dysregulated gene. Some of them such as chlorpromazine, nelfinavir, dasatinib, tamoxifen, and dabrafenib share more than one DEG identified in the course of testing them on the treatment of the

disease. Chlorpromazine, from among these drugs, has a therapeutic effect on the Middle East respiratory syndrome coronavirus (MERS-CoV) [35], nelfinavir has a therapeutic impact on human immunodeficiency viruses (HIV) and hepatitis C [36], and riboflavin on HIV. Although dabrafenib, gefitinib, and dasatinib have been approved in the treatment of various cancers, they affect the expression of genes that are dysregulated in viral diseases, especially in coronavirus-related ones. Niclosamide is also an FDA-approved anthelmintic drug which has the therapeutic potential in the coronavirus family [37]. Selamectin, which is a topical parasiticide and anthelmintic drug and has been approved for treating dogs and cats, was introduced as an efficient drug in the treatment of COVID-19 [38]. Since we did not observe any connection between the dysregulated genes and COVID-19 in CTD, we did not mention it in Fig. 7. A majority of our proposed drugs have many interactions with a number of viral diseases, so they have the therapeutic potential on COVID-19. Amodiaquine, vidarabine, and quinacrine are also associated with a large number of other viral diseases, including the coronavirus family. The complete information on the proposed drugs can be found in the Supplementary file 4.


Fig. 7 Candidate Drugs and their relationship with COVID-19
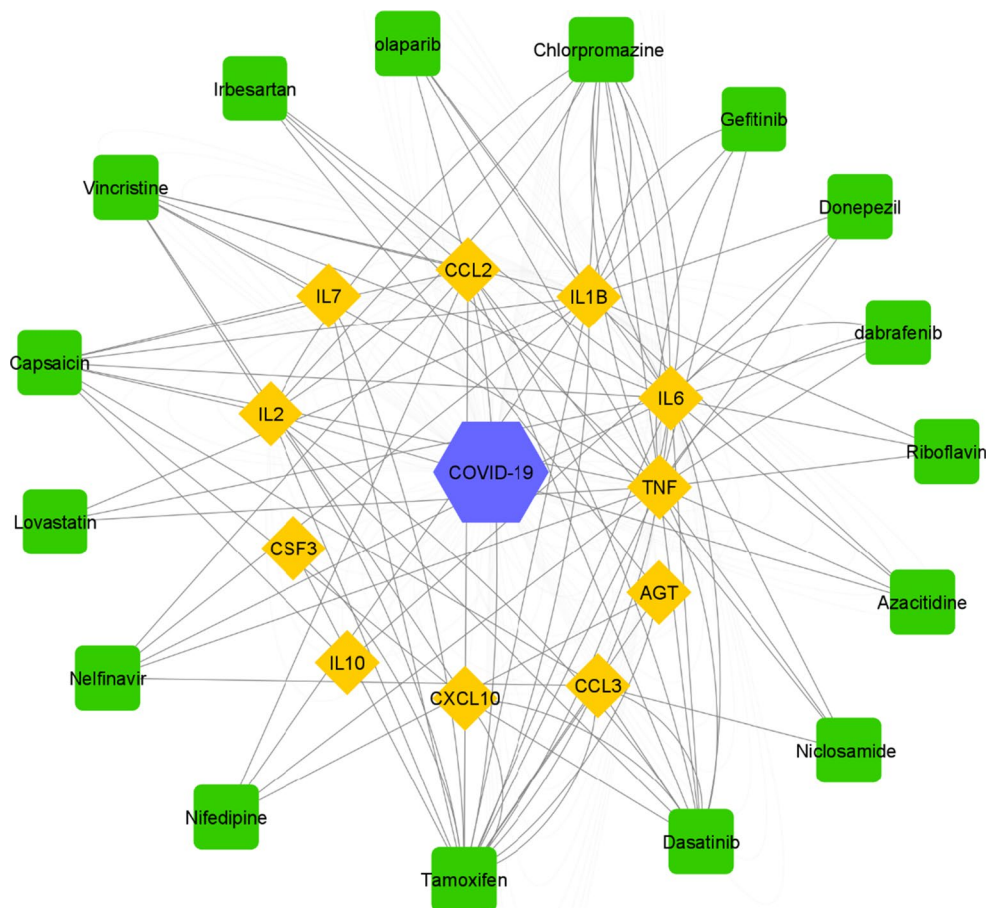
Chemical structural similarities between drugs

To assess the chemical structures of the candidate drugs, we investigated the similarities among the chemical structures of the proposed drugs and hydroxychloroquine, lopinavir,

disulfiram, and nitazoxanide [39]. As mentioned earlier, we examined a wide range of small molecules in the drug bank database. Upon examination, we found 45 small molecules that could be chosen as promising drugs for treating COVID-19. From among these molecules, some representatives,
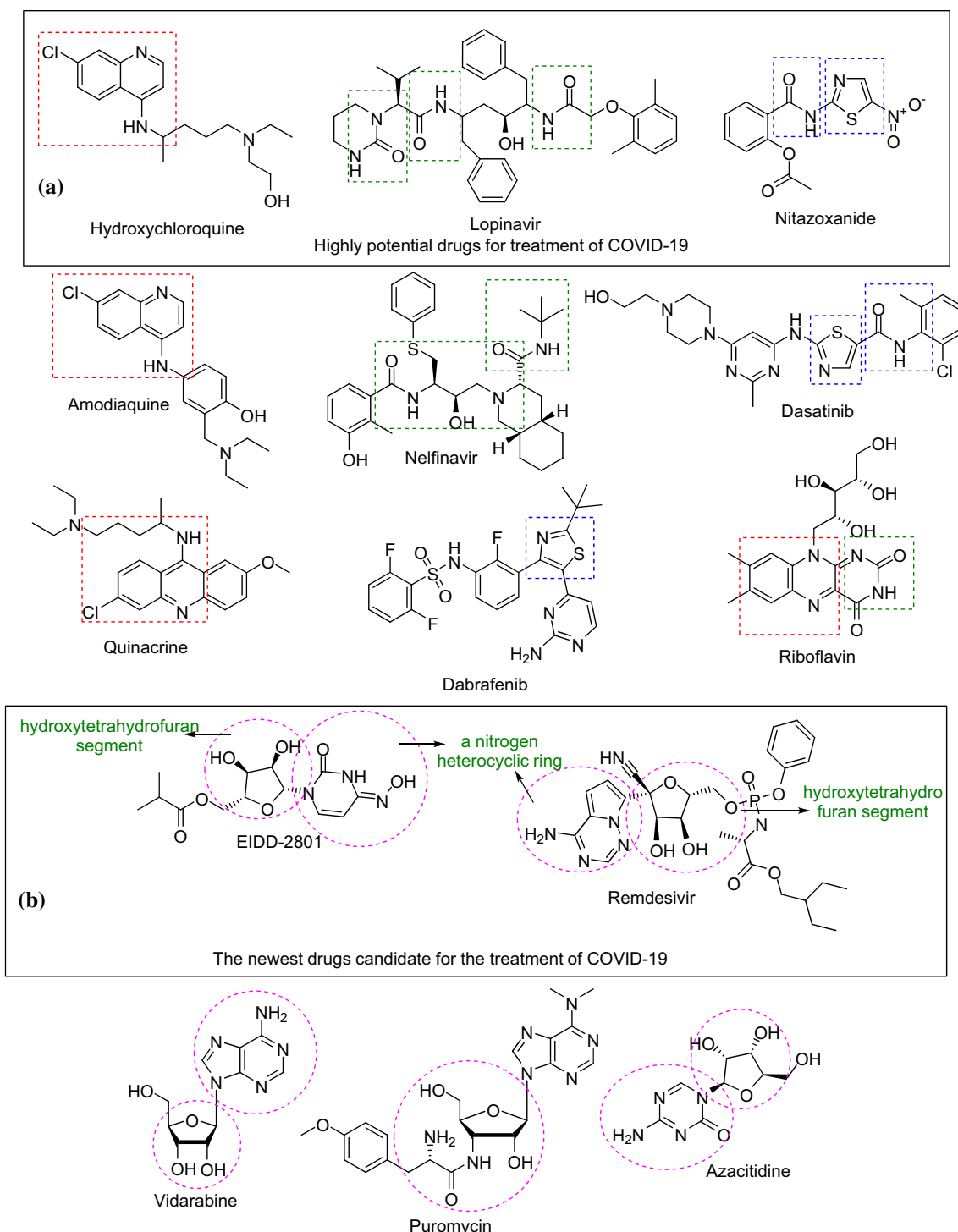
Fig. 8 Chemical structural comparison between the best small molecules, which have the potential to be used as a drug, and the previously proposed drugs

which possess some features in common with the medicines that produce rather satisfactory results in protecting the body against this type of coronavirus, were identified. Hydroxychloroquine contains a nitrogen aromatic heterocyclic ring named quinoline; lopinavir contains amides functional groups, and nitazoxanide has a thiazole ring (Fig. 8, box A). The same kinds of features were identified in amodiaquine, nelfinavir, dabrafenib, and dasatinib, respectively. Moreover, it was found that vidarabine, puromycin, and azacitidine contain hydroxytetrahydrofuran resembling remdesivir as well as EIDD-2801 (Fig. 8, box B), which was suggested for treating COVID-19.

Proposing various drugs has made it possible for researchers and medical specialists to discover an effective as well as efficient remedy at a rapid pace. Considering both the devised and previously used drugs, herein, we rigorously concentrated on their chemical structures similarities. we aimed to devise modified small molecules which compared to the previous candidates have fewer side effects and employ a synthetic route to create the desired products easily and cost-effectively. Like hydroxychloroquine that comprises quinolones, ten drugs which contain a quinoline ring were yielded as the end results. Many of the quinoline derivatives, which are regarded as useful compounds, can be found in a number of biologically active compounds and medicines [40, 41]. To illustrate this, consider the fact that quinoline has several antimalarial derivatives such as chloroquine, quinine, and primaquine [42]. Our method proposes amodiaquine (used in the treatment of malaria) and quinacrine that is considered to be an effective medication in the treatment of COVID-19. Like hydroxychloroquine, amodiaquine and quinacrine both have a quinoline ring and aniline derivatives. To conduct a follow-up study, we found that like nitazoxanide which has a thiazole ring and is an effective remedy for COVID-19, dasatinib and dabrafenib contain the thiazole heterocyclic motif. The thiazole core, which has attracted widespread attention, is considered to be a pivotal compound in the field of drug discovery. This structure is found not only in a biologically active molecule like thiamine (vitamin B_1) but also in some fungicides such as tricyclazole as well as nonsteroidal anti-inflammatory drugs, namely meloxicam [43, 44]. Besides, lopinavir, which contains two amides as well as some functional groups, is a pseudo peptide compound. Amides, which are considered to be the backbone of vital organic structures of nature, enzymes, and proteins in which peptide bonds play a significant role in the human's body, are one of the most important functional groups in biochemistry [45]. From among the proteins, both the di- and tripeptides from our pseudo peptides display a wide range of noticeable bioactivities such as antidiabetic, antibacterial, and antitumor activities [46]. The machine has introduced some drug structures which significantly contain pseudo peptide scaffolds.

As evidenced in Fig. 8, nelfinavir, which is an antiretroviral drug and is used in the treatment of HIV, has been selected by our approach. This drug contains some functional groups that are similar to those found in Lopinavir. Furthermore, riboflavin which is known as vitamin B_2 and has approximately the same structure (e.g., benzopyrazine and cyclic urea) as hydroxychloroquine and lopinavir do, was chosen by the machine as a remedy for COVID-19.

Riboflavin, which is used as a dietary supplement, is a vitamin found in food. Recent studies argue that riboflavin might prove useful in treating viral diseases [47].

It worth noting that a vast number of scientific studies have introduced a number of drugs and small molecules to treat COVID-19. Remdesivir, which has been previously suggested as a potential treatment for Ebola (in the literature, 2016), is also being researched as a possible remedy for COVID-19 [48–50]. Besides, EIDD-2801, as an experimental small molecule, was found to possess some therapeutic potential to treat COVID-19 [51]. Remdesivir and EIDD-2801 both contain hydroxytetrahydrofuran in which at least one nitrogen heterocyclic ring is connected to the central core. Through conducting a precision chemical analysis of the proposed small molecules, the machine firmly showed that some drug-containing hydroxy tetrahydrofurans and a heterocyclic nitrogen compound might prove to be effective in treating COVID-19. Interestingly enough, these small molecules skeletally resemble EIDD-2801 and Remdesivir, and in contrast to EIDD-2801 which is considered to be an experimental small molecule, these specified compounds are regarded as a drug. For example, vidarabine, which fits into the category of hydroxychloroquine, resembles EIDD and possesses a hydroxy tetrahydrofuran core with a heterocyclic nitrogen compound.

Conclusion

In this study, a multimodal restricted Boltzmann machine (mm-RBM) was employed to cluster two types of drug data: a fingerprint and DEGs. The first type of data was a binary data showing the chemical structures. The second one was extracted from drug-induced perturbations in cell lines.

In the proposed multimodal RBM model, first, the intrinsic correlations within each input modality were encoded using the modality-specific hidden variables. Next, the intra-modality features were fused by merging unknown variables, and a typical representation of cross-platform features was formed. The proposed approach indicates that data integration yields significant clusters. Therefore, the clusters consisting of drugs used for curing COVID-19 were chosen to discover medications which may prove useful in treating the disease. The introduced drugs, which have antiviral properties, are similar to advanced drugs that have been applied to

control COVID-19. Although the outcomes are significant and seem to yield a satisfactory explanation for the recent pandemic, further clinical research such as in vitro or in vivo tests need to be carried out.

Compliance with ethical standards

Conflict of interest The authors have no relevant conflicts of interest to disclose.

References

1. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, Doig A, Guilliams T, Latimer J, McNamee C (2019) Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov 18(1):41–58. https://doi.org/10.1038/nrd.2018.168

2. Athauda D, Foltynie T (2018) Drug repurposing in Parkinson's disease. CNS Drugs 32(8):747–761. https://doi.org/10.1007/s40263-018-0548-y

3. Masoudi-Sobhanzadeh Y, Omidi Y, Amanlou M, Masoudi-Nejad A (2019) DrugR+: a comprehensive relational database for drug repurposing, combination therapy, and replacement therapy. Computers Biol Med 109:254–262. https://doi.org/10.1016/j.compbiomed.2019.05.006

4. Masoudi-Sobhanzadeh Y, Omidi Y, Amanlou M, Masoudi-Nejad A (2020) Drug databases and their contributions to drug repurposing. Genomics 112:1087–1095

5. Masoudi-Sobhanzadeh M-N (2020) Synthetic repurposing of drugs against hypertension: a datamining method based on association rules and a novel discrete algorithm. BMC Bioinform 21:313

6. Masoudi-Sobhanzadeh Y, Omidi Y, Amanlou M, Masoudi-Nejad A (2019) Trader as a new optimization algorithm predicts drug-target interactions efficiently. Sci Rep 9(1):1–14. https://doi.org/10.1038/s41598-019-45814-8

7. Dudley JT, Deshpande T, Butte AJ (2011) Exploiting drug–disease relationships for computational drug repositioning. Briefings Bioinform 12(4):303–311. https://doi.org/10.1093/bib/bbr013

8. Iorio F, Rittman T, Ge H, Menden M, Saez-Rodriguez J (2013) Transcriptional data: a new gateway to drug repositioning? Drug Discov Today 18(7–8):350–357. https://doi.org/10.1016/j.drudis.2012.07.014

9. Brown AS, Patel CJ (2017) MeSHDD: literature-based drug-drug similarity for drug repositioning. J Am Med Inform Assoc 24(3):614–618. https://doi.org/10.1093/jamia/ocw142

10. Xu Y, Dai Z, Chen F, Gao S, Pei J, Lai L (2015) Deep learning for drug-induced liver injury. J Chem Inform Modeling 55(10):2085–2093. https://doi.org/10.1021/acs.jcim.5b00238

11. Ramachandram D, Taylor GW (2017) Deep multimodal learning: a survey on recent advances and trends. IEEE Signal Process Magazine 34(6):96–108. https://doi.org/10.1109/msp.2017.2738401

12. Ngiam J, Khosla A, Kim M, Nam J, Lee H, Ng AY (2011) Multimodal deep learning. In: ICML.

13. Hooshmand SA, Jamalkandi SA, Alavi SM, Masoudi-Nejad A (2020) Distinguishing drug/non-drug-like small molecules in drug discovery using deep belief network. Mol Diversity. https://doi.org/10.1007/s11030-020-10065-7

14. Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, Ma'ayan A (2016) The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database. https://doi.org/10.1093/database/baw100

15. Koleti A, Terryn R, Stathias V, Chung C, Cooper DJ, Turner JP, Vidović D, Forlin M, Kelley TT, D'Urso A (2018) Data portal for the library of integrated network-based cellular signatures (LINCS) program: integrated access to diverse large-scale cellular perturbation response data. Nucleic Acids Res 46(D1):D558–D566. https://doi.org/10.1093/nar/gkx1063

16. Marlin B, Swersky K, Chen B, Freitas N (2010) Inductive principles for restricted Boltzmann machine learning. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics, pp. 509–516.

17. Wen M, Zhang Z, Niu S, Sha H, Yang R, Yun Y, Lu H (2017) Deep-learning-based drug–target interaction prediction. J Proteome Res 16(4):1401–1409. https://doi.org/10.1021/acs.jproteome.6b00618

18. Zeng X, Zhu S, Liu X, Zhou Y, Nussinov R, Cheng F (2019) DeepDR: a network-based deep learning approach to in silico drug repositioning. Bioinformatics 35(24):5191–5198. https://doi.org/10.1093/bioinformatics/btz418

19. Napolitano F, Zhao Y, Moreira VM, Tagliaferri R, Kere J, D'Amato M, Greco D (2013) Drug repositioning: a machine-learning approach through data integration. J Cheminform 5(1):30. https://doi.org/10.1186/1758-2946-5-30

20. Tan F, Yang R, Xu X, Chen X, Wang Y, Ma H, Liu X, Wu X, Chen Y, Liu LJMB (2014) Drug repositioning by applying 'expression profiles' generated by integrating chemical structure similarity and gene semantic similarity. Mol BioSyst 10(5):1126–1138. https://doi.org/10.1039/c3mb70554d

21. Suk HI, Lee SW, Shen D, AsDN I (2014) Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. NeuroImage 101:569–582. https://doi.org/10.1016/j.neuroimage.2014.06.077

22. Liang M, Li Z, Chen T, Zeng J (2014) Integrative data analysis of multi-platform cancer data with a multimodal deep learning approach. IEEE/ACM Trans Comput Biol Bioinform 12(4):928–937. https://doi.org/10.1109/tcbb.2014.2377729

23. Masoudi-Sobhanzadeh Y, Motieghader H, Masoudi-Nejad A (2019) FeatureSelect: a software for feature selection based on machine learning approaches. BMC Bioinform 20:170

24. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M Tensorflow: a system for large-scale machine learning. In: 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16), 2016. pp 265–283

25. Chollet F (2015) keras, GitHub. GitHub repository, https://github.com/fchollet/keras

26. Zeng H, Qiu C, Cui Q (2015) Drug-path: a database for drug-induced pathways. Database. https://doi.org/10.1093/database/bav061

27. Ngiam J, Khosla A, Kim M, Nam J, Lee H, Ng AJML (2011) BMultimodal deep learning. In: Proc. 28th int. conf. mach. learn

28. Andrews PR, Mark LC (1982) Structural specificity of barbiturates and related drugs. 57(4):314–320. https://doi.org/10.1097/00000542-198210000-00014

29. Gao J, Tian Z, Yang X (2020) Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19

associated pneumonia in clinical studies. Bioscience Trends 4(1):72–73. https://doi.org/10.5582/bst.2020.01047

30. Dong L, Hu S, Gao J (2020) Discovering drugs to treat corona-virus disease 2019 (COVID-19). Drug Discoveries Therapeutics 14(1):58–60. https://doi.org/10.5582/ddt.2020.01012

31. Gautret P, Lagier J-C, Parola P, Meddeb L, Mailhe M, Doudier B, Courjon J, Giordanengo V, Vieira VE, Dupont HT (2020) Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. Int J Antimicrobial Agents 56(1):105949. https://doi.org/10.1016/j.ijantimicag

32. Srivatsan Padmanabhan M (2020) Potential dual therapeutic approach against SARS-CoV-2/COVID-19 with Nitazoxanide and Hydroxychloroquine. https://doi.org/https://doi.org/10.13140/RG.2.2.28124.74882

33. Li G, De Clercq E (2020) Therapeutic options for the 2019 novel coronavirus (2019-nCoV). Nat Rev Drug Discov, 19 (3): 149–150 https://doi.org/10.1038/d41573-020-00016-0

34. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, McMorran R, Wiegers J, Wiegers TC, Mattingly CJ (2019) The compara-tive toxicogenomics database: update 2019. Nucleic Acids Res 47(D1):D948–D954. https://doi.org/10.1093/nar/gky868

35. de Wilde AH, Jochmans D, Posthuma CC, Zevenhoven-Dobbe JC, van Nieuwkoop S, Bestebroer TM, van den Hoogen BG, Neyts J, Snijder EJ (2014) Screening of an FDA-approved compound library identifies four small-molecule inhibitors of Middle East respiratory syndrome coronavirus replication in cell culture. Antimicrobial agents Chemother 58(8):4875–4884. https://doi.org/10.1128/AAC.03011-14

36. Toma S, Yamashiro T, Arakaki S, Shiroma J, Maeshiro T, Hibiya K, Sakamoto N, Kinjo F, Tateyama M, Fujita J (2009) Inhibition of intracellular hepatitis C virus replication by nelfinavir and syn-ergistic effect with interferon-α. J Viral Hepatitis 16(7):506–512. https://doi.org/10.1111/j.1365-2893.2009.01102.x

37. Xu J, Shi P-Y, Li H, Zhou J (2020) Broad spectrum antiviral agent niclosamide and its therapeutic potential. ACS Infectious Dis 6(5):909–915. https://doi.org/10.1021/acsinfecdis.0c00052

38. Fan H-H, Wang L-Q, Liu W-L, An X-P, Liu Z-D, He X-Q, Song L-H, Tong Y-G (2020) Repurposing of clinically approved drugs for treatment of coronavirus disease 2019 in a 2019-novel coro-navirus (2019-nCoV) related coronavirus model. Chin Med J 133(9):1051–1056. https://doi.org/10.1097/CM9.0000000000000797

39. Barlow A, Landolf KM, Barlow B, Yeung SYA, Heavner JJ, Claassen CW, Heavner MS (2020) Review of emerging pharma-cotherapy for the treatment of coronavirus disease 2019. Pharma-cotherapy: J Human Pharmacol Drug Therapy, 40(5): 416–437 https://doi.org/10.1002/phar.2398

40. Shang XF, Morris-Natschke SL, Liu YQ, Guo X, Xu XS, Goto M, Li JC, Yang GZ, Lee KH (2018) Biologically active quinoline and quinazoline alkaloids part I. Med Res Rev 38(3):775–828. https://doi.org/10.1002/med.21466

41. Jain S, Chandra V, Jain PK, Pathak K, Pathak D, Vaidya A (2016) Comprehensive review on current developments of quinoline-based anticancer agents. Arab J Chem 12(8):4920–4946. https://doi.org/10.1016/j.arabjc.2016.10.009

42. Paul A, Egan TJ, Ross DC (1996) The mechanism of action of quinolines and related anti-malarial drugs. South African J Sci 92(1):11–14

43. Chen L, Zhu Y-J, Fan Z-J, Guo X-F, Zhang Z-M, Xu J-H, Song Y-Q, Yurievich MY, Belskaya NP, Bakulev VA (2017) Synthesis of 1, 2, 3-thiadiazole and thiazole-based strobilurins as potent fungicide candidates. J Agric Food Chem 65(4):745–751. https://doi.org/10.1021/acs.jafc.6b05128

44. Obach RS, Kalgutkar AS, Ryder TF, Walker GS (2008) In vitro metabolism and covalent binding of enol-carboxamide deriva-tives and anti-inflammatory agents sudoxicam and meloxicam: insights into the hepatotoxicity of sudoxicam. Chem Res Toxicol 21(9):1890–1899. https://doi.org/10.1021/tx800185b

45. Sieńczyk M, Podgórski D, Błażejewska A, Kulbacka J, Saczko J, Oleksyszyn J (2011) Phosphonic pseudopeptides as human neutrophil elastase inhibitors—a combinatorial approach. Bioor-ganic Med Chem 19(3):1277–1284. https://doi.org/10.1016/j.bmc.2010.12.008

46. Liu J-Z, Song B-A, Fan H-T, Bhadury PS, Wan W-T, Yang S, Xu W, Wu J, Jin L-H, Wei X (2010) Synthesis and in vitro study of pseudo-peptide thioureas containing α-aminophosphonate moi-ety as potential antitumor agents. Eur J Med Chem 45(11):5108–5112. https://doi.org/10.1016/j.ejmech.2010.08.021

47. Bowers JM, Bert-Moreno A (2004) Treatment of HAART-induced lactic acidosis with B vitamin supplements. Nutrition Clin Prac-tice 19(4):375–378. https://doi.org/10.1177/0115426504019004375

48. Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, Shi Z, Hu Z, Zhong W, Xiao G (2020) Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell Res 30(3):269–271. https://doi.org/10.1038/s41422-020-0282-0

49. Al-Tawfiq JA, Al-Homoud AH, Memish ZA (2020) Remdesi-vir as a possible therapeutic option for the COVID-19. Travel medicine and infectious disease 34:101615–101615. https://doi.org/10.1016/j.tmaid.2020.101615

50. Masoudi-Sobhanzadeh Y (2020) Computational-based drug repur-posing methods in COVID-19. Bioimpacts 10: 205–206 https://doi.org/10.34172/bi.2020.25

51. Rothan HA, Byrareddy SN (2020) The epidemiology and patho-genesis of coronavirus disease (COVID-19) outbreak. J Autoim-munity 109:102433. https://doi.org/10.1016/j.jaut.2020.102433