International Journal Dental and Medical Sciences Research
Volume 3, Issue 1, Jan-Feb 2021 pp 871-882 www.ijdmsrjournal.com ISSN: 2582-6018

Artificial Intelligence Assisted Ab-Initio Modelling of Computational Drug Discovery for Covid-19

Kamal Jain,Ankush Rai
Indian Institute of Technology, Roorkee

Submitted: 02-02-2021 Revised: 15-02-2021 Accepted: 18-02-2021

ABSTARCT: The rapid pandemic spread of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)-associated coronavirus disease 2019 (COVID-19) is increasingly impacting the healthcare infrastructure, the economic situation, and socio-cultural interactions globally. In the current spread of novel coronavirus (SARS-CoV-2), antiviral drug discovery is of great importance. But discovering such an investigational drug is a tedious process with lots of involvement of trial and error which usually require lots of time, skilled labor and monetary resources. In this research we present an Artificial Intelligence (AI) based technique to generate new medicinal formulation for Covid-19 virus. We deploy a hybrid algorithm in combination with Ab-Initio modeling of chemical compounds which reduces the time and labor spent on formulation by resorting to age old trial and error basis.

Keyword: Covid-19, Computational Drug Design, Artificial Intelligence, Ab-Initio Modeling, Precession Medicine.

I. INTRODUCTION

The unprecedented challenges posed by the coronavirus disease 2019 (COVID-19) pandemic highlight the urgency for applying clinical pharmacology and drug development in (i) identifiable of drug and dosage optimization for COVID-19 therapies, (ii) approaching therapeutic dilemmas in clinical trial settings, and (iii) maximizing value of information from impacted non–COVID-19 trials. More than ever, we have a responsibility for adaptive evidence synthesis with a Totality of Evidence mindset in this race against time across biomedical research, clinical practice, drug development, and regulation. The AI can be used to computationally design a drug on computer and can be simulated to check its effectiveness against corona virus.Thankfully, there is reason to be optimistic that we can improve the drug discovery system with our in-house built artificial intelligence for drug discovery.

We should pay attention to development of new techniques of treating old diseases by drug repurposing. Drug repurposing will be accomplished by analyzing drug−drug association as well as drug−target associations. As verified by a survey on DrugBank results, on averaging every drug has 3 therapeutic strategies and each gene product has 4.7 drugs. This finding reveals that polypharmacology is a widespread condition. It is most critical job that ensures both the prescription medication and candidate drug can do little harm to patients. Polypharmacology will lead to rational assessment of potential new groups of medications which will potentially show less adverse effects. Computer-aided drug design (CADD) has made a significant contribution to contemporary drug research and production. Since high precision and efficiency are needed, a hierarchical system integrating various types of multi - objective optimization is implemented. A simple docking scoring feature, like Glide's, can screen a wide library quickly and efficiently, but it is not quite effective. However, scoring functions based on molecular mechanical field (MMFF) are physical and more precise but much less effective. Because of the ever growing computing capacity, molecular mechanics based free energy measurement methods, such as with the end point MM-PBSA (molecular mechanics Poisson−Boltzmann surface area) and MM-GBSA (molecular mechanics generalized Born surface area) and alchemical thermodynamic integration system (TI) have been widely implemented in structure-based drug development programs.

We also developed a dynamic drug formulation with verification by virtual screening (VS) based on binding energy to match the efficiency and accuracy and increase the success rate of reasonable drug design. In this work, We used ab-initio modeling techniques to provide means of targeting the SARS Coronavirus protease. A compact docking approach and an MBSA-weighted solvent-accessible surface area (WSAS) were added to elucidate the structure of SARS-CoV-2 main protease. Compared to the tests, CADD is more effective so it can provide care for a potential epidemic disease outbreak like COVID-

International Journal Dental and Medical Sciences Research
Volume 3, Issue 1, Jan-Feb 2021 pp 871-882 www.ijdmsrjournal.com ISSN: 2582-6018

19. The detailed ligand-residue structure and the decay of binding sites offers insight into developing inhibitors of the main protease of SARS CoV-2.

II. METHODOLOGY

1.1. Experimental Setup

The proposed algorithm is prototyped over Python 36 under Windows platform, with hardware specifications of Intel's third generation 8-core microprocessor, 16GB RAM giving the clocking speed of 2.7 GHz with 25GB Nvidia Quadro. The standardized databases used in the study are publicly available ZINC database, web scrapped PubChem Database& Drug Bank Database. Here, ZINC Database is a curated collection of commercially available chemical compounds prepared especially for virtual screening whereas PubChem is a database of chemical molecules and their activities against biological assays. The system is maintained by the National Center for Biotechnology Information; also The DrugBank database used in the study is a comprehensive, freely accessible, online database containing information on drugs and drug targets. As both a bioinformatics and a cheminformatics resource, DrugBank combines detailed drug data with comprehensive drug target information. (table 1) [22,23,24]. We have also used the newly released crystal structure of SARS-CoV-2 main protease and conducted novel drug formulation followed by multiscale drug repurposing screenings. Apart from drug database we have also used drug connectivity map [25]. The Connectivity Map, or also commonly referred as CMap, is used to train our AI for predicting the effect of drug composition on genetic pathways. Cmap is a resource that uses transcriptional expression data to probe relationships between diseases, cell physiology, and therapeutics. The changes in gene expression, or "signatures," that arise from a disease, genetic perturbation (knockdown or overexpression of a gene) or treatment with a small molecule are compared for similarity to all perturbational signatures in the database.

Table.1. List of database used in the test.

Database	Total Compounds Used in Study	Mode of Extraction	Availability of 3D Molecular Structure	Total Molecules in the Database
Zinc Database	65,000	Download Link Web Scrapped	No	7,27,842
PubChem Database	54,000	Download API	Yes	7,27,842
Drug Bank Database	5,000		Yes	>30,000

1.2. Algorithms for Automating Compound Formulation Using Ab-initio Modeling

For the development of AI based on co-simulated competitive neural network to generate a list of new medicines along with the discovery of existing medicines. Each living cells are made up of sequence of protein compounds. This, protein compounds in turn combine together to form cells, these cells combined to form tissue and tissues combined to form organs. Hence, when a virus attacks a living organism it changes the sequence of this protein to misbalance their normal regulatory function. In turn it makes our own cells behave against its own like that of a computer virus and it gets on replicating and disrupting regulatory mechanism of other cells on and on. We can simulate a drug or compound which can combine with these viruses to disrupt its own regulatory mechanism and thereby disallowing it to bond with our cells.

International Journal Dental and Medical Sciences Research
Volume 3, Issue 1, Jan-Feb 2021 pp 871-882 www.ijdmsrjournal.com ISSN: 2582-6018

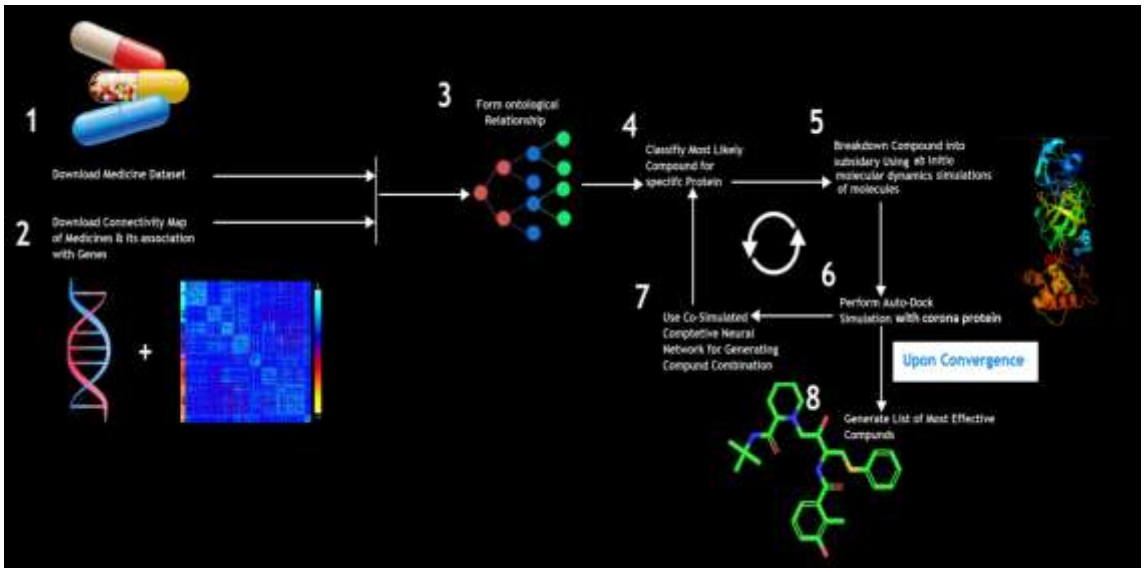

Figure 2: The above flowchart illustrates Our process for computational drug discovery.

Our AI system first formulates the connection between protease and existing chemical compound to form a logical connection of association between protein and disease (figure 1). Upon which each of this most likely candidate identified in the first phase is broken down into its subsidiary and more stable components. For evaluation, we read two paired Instance of chemical compounds with drugs and its connectivity map with respect to genes in a computational Workspace W (b) Instance of durg end U, Action SetsAS_i and Matrix Model of Tree of Actions M_x, L_U&L_W which are the set of optimized binding levels from U & W respectively.Compute the Pointing Correlation state P as:

$$P = \frac{1}{L_N} \sum_{p_i}^{L_W - 1} \left[\sum_{p_2}^{L_U - 1} S_{p_1,p_2}(t_i, f_1, f_2) \right] \left[\sum_{p_2}^{L_U - 1} S'_{p_1,p_2}(t_i, f_1, f_2) \right]$$

where, L_N are the universal set of level for the drugs, p_i&p_2 are the adjoincmap associated drugs intersection with the levels L_W&L_U respectively, S_{p_1,p_2}&S'_{p_1,p_2} are the sets of drugs constraint layout for the porteasemolecule positioning with its ab-initio evaluated energy model saved in levels and between its intersection of adjoint pixels and the superpositioned drug-proteaseenergy density layout (evaluated from SIESTA an ab-initio module) of differing state at the current instance of the drug U [28]. Also, t_i is the collection of patterns for the weighted superposed state P_c(initially its value is set to 0), f_1, f_2 are the two sub-module delay drug units with a minimal time delayst_i. Later we calculate tree of Action based on continuous feedback loop generated from previous process. Now, the weighted values of drug protease association with action sets can be summed as:

International Journal Dental and Medical Sciences Research
Volume 3, Issue 1, Jan-Feb 2021 pp 871-882 www.ijdmsrjournal.com ISSN: 2582-6018

$$M_x = \begin{pmatrix} t_1 \begin{bmatrix} P_1 \\ P_4 \\ P_8 \end{bmatrix} = AS_1 \\ t_2 \begin{bmatrix} P_3 \\ P_9 \\ P_6 \end{bmatrix} = AS_2 \\ t_3 \begin{bmatrix} P_2 \\ P_5 \\ P_7 \end{bmatrix} = AS_3 \\ \vdots \\ t_i \begin{bmatrix} P_0 \\ P_5 \\ P_c \end{bmatrix} = AS_i \end{pmatrix}$$

where, AS_i is the automated classified action sets. The data collected from previous process is latter used in ab-intio modeling from ground up i.e, atomic simulation is performed to build the compound with most affinity to bind with protein of corona virus and inhibit it to bind with human cells. As the AI formulates new formulae of the chemical compound against corona virus it test for its binding affinity and thereby sort the best candidates on its own. Now, we evaluate a policy to reinforce the policy sets for generation of novel medicine by synchronously sequencing the successful pairs of molecular clumping. This sets of data stream is latter feed to Recurrent Neural Network (RNN) for implementation for vector routing of weights is formalized as follows: Given a sequence of input vectors (X_1,X_2,X_3, ...,X_n) of drug-gene sharing, the RNN computes a sequence of hidden states (H_1,H_2,H_3, ...,H_n) of pointers from the chunk table, and a sequence of outputs (O_1,O_2,O_3, ...,O_n) is encoded, by iterating the following equations

for 1: b_{jq}
for t = 1 to X:

$$H_n = \tanh(w_{HX} X_N + w_{HH} H_{N-1} + B_N) \quad (1)$$

$$k_i = w_{HH} . X_N + B_N + \sum_{i+1}^{N} \tanh(\delta_H . (1 - y_{N-1}) \quad (2)$$

while j<X

$$k_j = w_{HO} . X_N + \sum_{i+1}^{N} \tanh(\delta_O . y_{N-1}) \quad (4)$$

$$O_N = w_{OH} H_N + B_O \quad (5)$$

end loop

$$\Delta w_{XX} = \eta . \delta_O . k_i . O_N \quad (6)$$

end while loop

$$PCC = \prod_{i=1}^{N} \Delta w_{XX}\, t_{f_t}(R_i, L_i) \quad (5)$$

end loop

In these equations, w_{HX} is the input-to-hidden weight matrix, w_{HH} is the hidden-to-hidden (or recurrent) weight matrix, w_{OH} is the hidden-to-output weight matrix, and the vectors B_N and B_O are the biases. The expression replaces the inputs received form feedback loops with a special initial bias vector checked for nonlinearity while ensuring that the training is done coordinate wise. η is the learning rate and k_i is the local induced field of activation potential for the i^{th} neuron, k_j is the co-activation neuron field for the next sequence of activation units,δ_H&δ_O are the pointer variable for the field & sub-field trace of a ip packet, respectively. Where the integrated co-routing involvement with the packet sharing & subsequent acknowledgment states derived from this test suite

International Journal Dental and Medical Sciences Research
Volume 3, Issue 1, Jan-Feb 2021 pp 871-882 www.ijdmsrjournal.com ISSN: 2582-6018

is depicted in table 2. The final list of generated molecules with its ADMET (Absorption, Distribution, Metabolism, Excretion and Toxicity) analysis is shown in Figure 2, table 3(A) & 3(B).

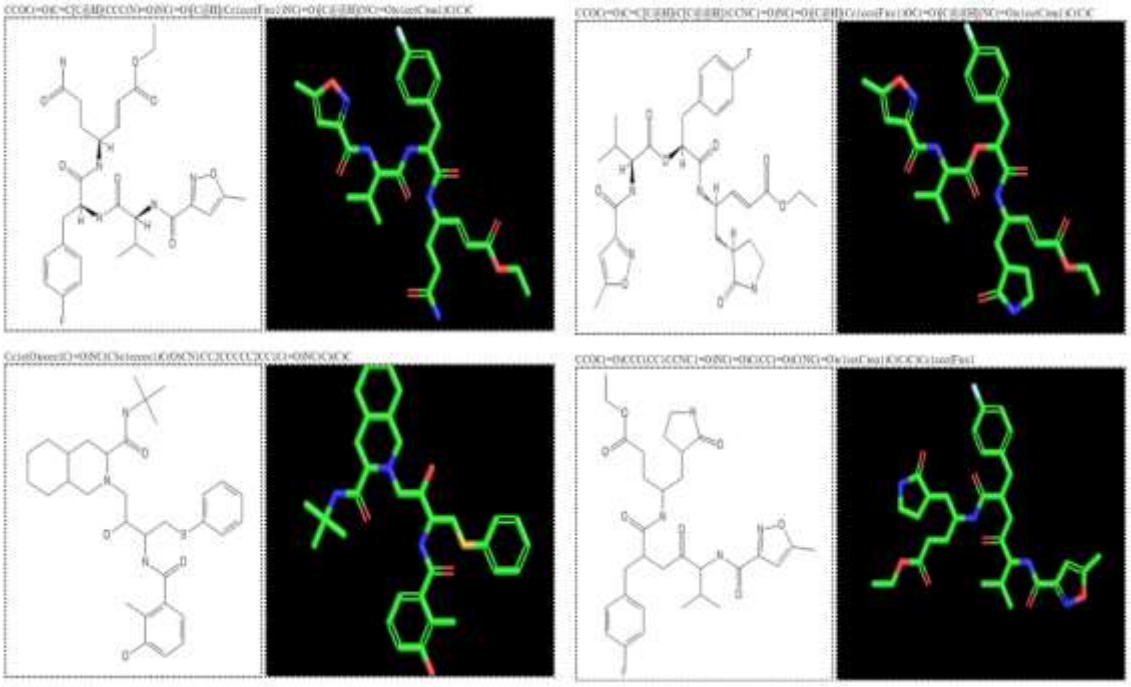

Figure 2: Structure of some of the generated medicinal formulation.

III. RESULTS

In our results we have identified around 70 compounds best suited against corona virus; out of which 19 are existing drugs. Interestingly, we have found in our study that mainly drugs used for HIV Inhibition, ant-malarial, anti-tumor drugs, and drugs for neurodegenerative disorders, arthrosis are most likely candidates to be found effective against corona virus in our simulation. We have not only be able to found novel drug formulae against corona virus but also few anti-HIV, ant-tumor, anti-malarial, anti-arthritis drugs and handful of ant-inflammatory drugs in the process. We hope our study can be used by ICMR and other medical practitioner and pharma companies to aid the nation in fighting against this pandemic. Widespread COVID-19 demands several prescription plans with early diagnosis as quickly as possible. Probably, such studies on repurposed drugs will demonstrate the promise of computational science. But for a successful care alternative is open with this AI generated medicine in a limited period of time. Although, it takes a lot of time to complete each data processing job for precision medicine but it reduces the time invested in trial and error basis. Furthermore, since performing docking screening on of all publicly available drugs database is not feasible. Thus, if we resort to use AI for precision medicine generation by strategic simplification & elimination from thousands of medicines by the utilization of AI over ab-initio modeling. The ab-initio computer simulations using SIESTA interface allows us to testformulation on atomic scale basis.The sizes of proteases to check for binding energy with Covid-19 virus are relatively high and there computational resources for the screening period for the proteases could be longer.

The takeaway from the table 2 is that Most of Antipicornaviral drugs are well suited to fight covid19 along with anti-rhinovirus drugs and HIV inhibition drugs and drugs to improve pharmacokinetics. Our study has been supported by the clinical results of dexamethasone & remdesivir from different research group across the world. As shown in Table 2 & 3 Compounds with binding energy greater than -5.1 is to considered effective against corona virus unlike anti-HIV compounds (like Lopinavir) or cocktail of medicines (like that of hydroxychloroquine, an antibiotic and zinc).Our newly generated compound by AI is giving binding emery with covid-19 to be ranging between -5 to -8 which is significant twice and in some cases thrice the greater amount than that of existing drugs like dexamethasone & remdesivir. Remdesivir with binding energy with covid19 protein ranges around -5.24 and it takes 11 days since infection to show

International Journal Dental and Medical Sciences Research
Volume 3, Issue 1, Jan-Feb 2021 pp 871-882 www.ijdmsrjournal.com ISSN: 2582-6018

recovery signs in patients. Hence, if we have novel compounds which have higher binding energy that of -5.24 are more likely to bond with covid19 and readily prove to be effective and can lead to faster recovery in less than 11 days. We have not only be able to found 25 novel drug compounds against corona virus but also few novel anti-HIV, ant-tumor, anti-malarial, anti-arthritis drugs and handful of ant-inflammatory drugs in the process.

Table 2: AI generated Drug Formulation

Computationally Designed Compound Against Covid19	**Common Name**	**Binding Energy**	**Properties and Usage**
CCOC(=O)C=C[C@H](CCC(N)=O)NC(=O)[C@H](Cc1ccc(F)cc1)NC(=O)[C@@H](NC(=O)c1cc(C)on1)C(C)C	Unknown	-8.37778	rhinovirus serotypes of antiviral group
Cc1c(O)cccc1C(=O)NC(CSc1ccccc1)C(O)CN1CC2CCCCC2CC1C(=O)NC(C)(C)C	Unknown	-6.67778	improving pharmacokinetics
CCOC(=O)C=C[C@H](C[C@@H]1CCNC1=O)NC(=O)[C@H](Cc1ccc(F)cc1)OC(=O)[C@@H](NC(=O)c1cc(C)on1)C(C)C	Unknown	-6.37778	Antipicornaviral compound
CCOC(=O)CCC(CC1CCNC1=O)NC(=O)C(CC(=O)C(NC(=O)c1cc(C)on1)C(C)C)Cc1ccc(F)cc1	Unknown	-6.35556	regulatable biocircuit systems:tunable protein expression
CCOP(=O)(COc1ccc(CC(NC(=O)OC2COC3OCCC23)C(O)CN(CC(C)C)S(=O)(=O)c2ccc(OC)cc2)cc1)OCC	Unknown	-6.2	Unknown
O=C(NC1CC2CCCC(C1)N2CCc1ccccc1)C1CC2CC(C1)C2C(=O)NC(Cc1ccccc1)C(=O)NC(Cc1ccccc1)C(=O)NC(Cc1ccccc1)C(=O)O	Unknown	-6.18889	poential for drug delivery TRANSFECTION OF NUCLEIC ACIDS, Zwitterionic reagents, boosting the performance of performance of chemiluminescent immunoassays,HIV-1 protease dimerization inhibitors
COC(=O)NC(C(=O)NC(Cc1ccccc1)C(O)CN(Cc1ccc(-c2ccccn2)cc1)NC(=O)C(NC(=O)OC)C(C)(C)C)C(C)(C)C	Unknown	-6.17778	COMPOUNDS AND METHODS FOR TREATING PAIN
CC(C)(C)NC(=O)C1CN(Cc2cccnc2)CCN1CC(O)CC(Cc1ccccc1)C(=O)NC1c2ccccc2CC1O	Unknown	-6.13333	cancer treatment: immunomodulator and anti-pathogenic agent
COC(=O)NC(C(=O)NCCCCC(CO)N(CC(C)C)S(=O)(=O)c1ccc(N)cc1)C(c1ccccc1)c1ccccc1	Unknown	-5.84444	HIV aspartyl protease inhibitors
CCOC(=O)C=C[C@H](C	Unknown	-5.73333	antipicornaviral agent

CC(N)=O)NC(=O)[C@H](Cc1ccc(F)cc1)OC(=O)[C@@H](NC(=O)c1cc(C)on1)C(C)C			
CC(C)CN(CC(O)C(Cc1ccccc1)NC(=O)OC1COC2OCCC12)S(=O)(=O)c1ccc2nc(NC3CCN(C4CCCC4)CC3)sc2c1	Unknown	-5.71111	BOOSTERS OF ANTIVIRALS
CC(C)(C)NC(=O)C1CC2CCCCC2CN1CC(O)C(Cc1ccccc1)NC(=O)C(CC(N)=O)NC(=O)c1ccc2ccccc2n1	Sakvinavir	-5.63333	immunomodulators
O=C1Nc2ccc(Cl)cc2C(C#CC2CC2)(C(F)(F)F)O1	Unknown	-5.6	hiv inhibitor
CCOC(=O)CCC(CC1CCNC1=O)NC(=O)C(CC(=O)C(NC(=O)c1cc(C)on1)C(C)C)Cc1ccc(F)cc1	Unknown	-5.57778	regulatable biocircuit systems:tunable protein expression
CCOC(=O)C=C[C@H](CCC(N)=O)NC(=O)[C@H](Cc1ccccc1)n1ccc(C)c(NC(=O)c2cc(C)on2)c1=O	Unknown	-5.52222	Inhibitors of severe acute respiratory syndrome
O=C(NC1CC2CCCC(C1)N2CCc1ccccc1)C1COC2CC(C1)C2C(=O)NC(Cc1ccccc1)C(=O)NC(Cc1ccccc1)C(=O)NC(Cc1ccccc1)C(=O)O	Unknown	-5.46667	Unknown
CC(C)(C)NC(=O)C1CC2CCC(CC1C(=O)NC(Cc1ccccc1)C(=O)O)C2C(=O)NC(Cc1ccccc1)C(=O)NC(Cc1ccccc1)C(=O)O	Unknown	-5.43333	probable HIV inhibiter
O=C(NCC1(c2ccccn2)CC1)NC1CC2CC(C1)C2C(=O)NC(Cc1ccccc1)C(=O)NC(Cc1ccccc1)C(=O)NC(Cc1ccccc1)C(=O)O	Unknown	-5.35556	probable for joint disorders, e.g. arthritis, arthrosis
CC(C)CN(CC(O)C(Cc1ccccc1)NC(=O)OC1COC2OCCC12)S(=O)(=O)c1ccc(N)cc1	Prezista	-5.32222	hiv inhibitor and booster to increase darunavir
CCCC1(CCc2ccccc2)CC(O)=C(C(CC)c2cccc(NS(=O)(=O)c3ccc(C(F)(F)F)cn3)c2)C(=O)O1	Unknown	-5.28889	hiv inhibitor
CCC(CC)COC(=O)C(C)NP(=O)(OCC1C(C(C(O1)(C#N)C2=CC=C3N2N=CN=C3N)O)O)OC4=CC=CC=C4	**remdesivir**	-5.24444	**Proven Effective against Covid-19**
Cc1cc(C(=O)Nc2cccn([C@@H](Cc3ccc(F)c(F)c3)	Unknown	-5.22222	Effective against SARS

International Journal Dental and Medical Sciences Research
Volume 3, Issue 1, Jan-Feb 2021 pp 871-882 www.ijdmsrjournal.com ISSN: 2582-6018

C(=O)N[C@H](C=CC(=O)OC(C)C)C[C@@H]3CCNC3=O)c2=O)no1			
O=C(NC1CC2CCCC21)C1CC2CC(C1)C2C(=O)NC(Cc1ccccc1)C(=O)NC(Cc1ccccc1)C(=O)NC(Cc1ccccc1)C(=O)O	Unknown	-5.2	potential novel antimicrobial
O=C(NC(Cc1ccccc1)C(=O)NC(Cc1ccccc1)C(=O)O)C1CC2CCC(C1)C2C(=O)NC(Cc1ccccc1)C(=O)NC(Cc1ccccc1)C(=O)O	Unknown	-5.17778	Apprimate for Zwitterionic reagents: improving the performance of chemiluminescent immunoassays

Table 2: Atomic Properties of Generated Drug Formulae

Formula	MW	#Heavy atoms	#Aromatic heavy atoms	Fraction Csp3	#Rotatable bonds	#H-bond acceptors	#H-bond donors
C28H36FN5O7	573.61	41	11	0.43	19	9	4
C32H45N3O4S	567.78	40	12	0.56	12	5	4
C30H37FN4O8	600.64	43	11	0.47	18	10	3
C31H41FN4O7	600.68	43	11	0.55	19	9	3
C33H49N2O12PS	728.79	49	12	0.61	21	13	2
C52H61N5O6	852.07	63	24	0.44	22	7	5
C38H52N6O7	704.86	51	18	0.45	22	9	5
C36H47N5O4	613.79	45	18	0.47	14	7	4
C33H44N4O6S	624.79	44	18	0.39	19	7	4
C28H35FN4O8	574.6	41	11	0.43	19	10	3
C38H53N5O7S2	755.99	52	15	0.63	16	10	3
C38H50N6O5	670.84	49	16	0.5	16	7	5
C14H9ClF3NO2	315.67	21	6	0.36	1	5	1
C31H41FN4O7	600.68	43	11	0.55	19	9	3
C29H33N5O7	563.6	41	17	0.31	16	8	3
C52H61N5O7	868.07	64	24	0.44	22	8	5
C43H52N4O8	752.89	55	18	0.44	20	8	6
C45H50N6O6	770.92	57	24	0.38	22	7	6
C27H37N3O7S	547.66	38	12	0.52	13	8	3
C31H33F3N2O5S	602.66	42	18	0.35	12	9	2
C27H35N6O8P	602.58	42	15	0.48	14	12	4
C31H33F2N5O7	625.62	45	17	0.35	15	10	3
C43H50N4O6	718.88	53	18	0.47	19	6	5
C46H50N4O8	786.91	58	24	0.35	22	8	6

Table 3(A): ADMET Analysis of generated Formulation

International Journal Dental and Medical Sciences Research
Volume 3, Issue 1, Jan-Feb 2021 pp 871-882 www.ijdmsrjournal.com ISSN: 2582-6018

Formula	MR	TPSA	iLOGP	XLOGP3	WLOGP	MLOGP	Silicos-IT Log P	Consensus Log P	ESOL Log S	ESOL Solubility (mg/ml)	ESOL Solubility (mol/l)	ESOL Class	Ali Log S	Ali Solubility (mg/ml)	Ali Solubility (mol/l)
C28H36FN5O7	145.54	182.72	3.5	2.28	1.89	0.59	3.75	2.4	-3.78	9.58E-02	1.67E-04	Soluble	-5.75	1.01E-03	1.76E-06
C32H45N3O4S	166.17	127.2	3.87	5.67	4.37	3.2	4.56	4.33	-6.36	2.47E-04	4.34E-07	Poorly soluble	-8.11	4.45E-06	7.83E-09
C30H37FN4O8	155.34	165.93	4.47	3.42	2.2	1.38	4.61	3.22	-4.72	1.14E-02	1.91E-05	Moderately soluble	-6.58	1.56E-04	2.60E-07
C31H41FN4O7	159.53	156.7	4.01	3.16	3.09	1.58	5.65	3.5	-4.49	1.94E-02	3.23E-05	Moderately soluble	-6.12	4.55E-04	7.57E-07
C33H49N2O12PS	180.84	186.58	5.14	3.64	5.48	0.73	2.37	3.47	-5.45	2.60E-03	3.57E-06	Moderately soluble	-7.25	4.13E-05	5.67E-08
C52H61N5O6	246.79	156.94	5.64	5.61	5.28	3.65	6.68	5.37	-7.49	2.78E-05	3.26E-08	Poorly soluble	-8.67	1.83E-06	2.15E-09
C38H52N6O7	193.76	171.22	3.56	5.6	4.06	1.76	4.11	3.82	-6.55	2.00E-04	2.84E-07	Poorly soluble	-8.96	7.77E-07	1.10E-09
C36H47N5O4	182.62	118.03	3.95	2.92	1.63	1.33	3.97	2.76	-4.86	8.53E-03	1.39E-05	Moderately soluble	-5.06	5.35E-03	8.71E-06
C33H44N4O6S	171.75	159.44	3.99	4.68	5.21	2.43	3.7	4	-5.71	1.22E-03	1.95E-06	Moderately soluble	-7.76	1.10E-05	1.75E-08
C28H35FN4O8	143.83	179.92	4	2.85	2.32	0.99	4.08	2.85	-4.14	4.14E-02	7.20E-05	Moderately soluble	-6.29	2.97E-04	5.16E-07
C38H53N5O7S2	206.13	179.18	4.97	6.26	5.73	2.5	4.05	4.7	-7.63	1.78E-05	2.35E-08	Poorly soluble	-9.81	1.17E-07	1.55E-10
C38H50N6O5	192.87	166.75	3.66	4.24	2.71	1.4	3.84	3.17	-5.86	9.34E-04	1.39E-06	Moderately soluble	-7.45	2.36E-05	3.53E-08
C14H9ClF3NO2	73.18	38.33	2.68	4.01	4.67	3.61	4.01	3.8	-4.47	1.07E-02	3.40E-05	Moderately soluble	-4.52	9.59E-03	3.04E-05
C31H41FN4O7	159.53	156.7	4.01	3.16	3.09	1.58	5.65	3.5	-4.49	1.94E-02	3.23E-05	Moderately soluble	-6.12	4.55E-04	7.57E-07
C29H33N5O7	150.26	175.62	3.03	1.62	2.17	0.88	3.49	2.24	-3.61	1.40E-01	2.48E-04	Soluble	-4.92	6.77E-03	1.20E-05
C52H61N5O7	247.88	166.17	4.9	4.73	4.66	2.9	6.26	4.69	-7.03	8.15E-05	9.39E-08	Poorly soluble	-7.95	9.76E-06	1.12E-08
C43H52N4O8	207.11	191	2.99	5.43	3.92	2.41	4.89	3.93	-6.85	1.06E-04	1.41E-07	Poorly soluble	-9.2	4.78E-07	6.35E-10
C45H50N6O6	214.12	178.62	4.35	5.51	4.03	2.6	5.43	4.38	-6.95	8.64E-05	1.12E-07	Poorly soluble	-9.02	7.36E-07	9.55E-10
C27H37N3O7S	142.2	148.8	3.2	2.94	3.46	1.18	1.46	2.45	-4.46	1.88E-02	3.44E-05	Moderately soluble	-5.73	1.03E-03	1.88E-06
C31H33F3N2O5S	153.8	113.97	3.68	6.97	9.37	3.74	6.55	6.06	-7.49	1.94E-05	3.22E-08	Poorly soluble	-9.18	4.01E-07	6.65E-10
C27H35N6O8P	150.43	213.36	3.24	1.91	2.21	0.18	-0.05	1.5	-4.12	4.58E-02	7.59E-05	Moderately soluble	-6.01	5.84E-04	9.69E-07
C31H33F2N5O7	161.52	161.63	3.81	2.9	3.25	1.99	4.58	3.31	-4.84	9.14E-03	1.46E-05	Moderately soluble	-5.95	6.94E-04	1.11E-06
C43H50N4O6	201.07	153.7	4.09	6.29	4.22	3.18	5.34	4.62	-7.26	3.98E-05	5.53E-08	Poorly soluble	-9.31	3.55E-07	4.94E-10
C46H50N4O8	217.14	191	3.13	6.49	4.12	2.7	5.7	4.43	-7.66	1.71E-05	2.18E-08	Poorly soluble	-10.3	3.97E-08	5.05E-11

Table 3(B): ADMET Analysis of generated Formulation

Formula	Silicos-IT class	GI absorption	BBB permeant	Pgp substrate	CYP 1A2 inhibitor	CYP2 C19 inhibitor	CYP 2C9 inhibitor	CYP 2D6 inhibitor	CYP 3A4 inhibitor	Synthetic Accessibility
C28H36FN5O7	Poorly soluble	Low	No	Yes	No	No	No	No	Yes	5.31
C32H45N3O4S	Poorly soluble	Low	No	Yes	No	Yes	No	No	Yes	5.58
C30H37FN4O8	Poorly soluble	Low	No	Yes	No	No	No	No	Yes	5.68
C31H41FN4O7	Poorly soluble	Low	No	Yes	No	No	No	No	Yes	5.54
C33H49N2O12PS	Poorly soluble	Low	No	Yes	No	Yes	No	No	Yes	6.9
C52H61N5O6	Insoluble	Low	No	Yes	No	No	No	No	Yes	8.25
C38H52N6O7	Poorly soluble	Low	No	Yes	No	No	No	No	Yes	6.24
C36H47N5O4	Poorly soluble	High	No	Yes	No	No	No	No	No	5.6
C33H44N4O6S	Poorly soluble	Low	No	Yes	No	No	Yes	No	No	5.39
C28H35FN4O8	Poorly soluble	Low	No	Yes	No	No	No	Yes	Yes	5.38
C38H53N5O7S2	Poorly soluble	Low	No	Yes	No	No	No	No	Yes	7.29
C38H50N6O5	Poorly soluble	Low	No	Yes	No	No	No	No	Yes	5.94
C14H9ClF3N	Modera	High	Yes	Yes	Yes	Yes	Yes	No	No	3.56

O2	tely soluble									
C31H41FN4O7	Poorly soluble	Low	No	Yes	No	No	No	No	Yes	5.54
C29H33N5O7	Poorly soluble	Low	No	Yes	No	No	Yes	No	Yes	5.18
C52H61N5O7	Insoluble	Low	No	Yes	No	No	No	No	Yes	8.56
C43H52N4O8	Poorly soluble	Low	No	Yes	No	No	No	No	Yes	6.99
C45H50N6O6	Insoluble	Low	No	Yes	No	No	Yes	No	Yes	7.16
C27H37N3O7S	Moderately soluble	Low	No	Yes	No	No	No	No	Yes	5.67
C31H33F3N2O5S	Insoluble	Low	No	Yes	No	Yes	No	Yes	Yes	5.29
C27H35N6O8P	Moderately soluble	Low	No	Yes	No	No	No	No	Yes	6.33
C31H33F2N5O7	Poorly soluble	Low	No	Yes	No	No	Yes	Yes	Yes	5.56
C43H50N4O6	Poorly soluble	Low	No	Yes	No	No	Yes	No	Yes	6.97
C46H50N4O8	Insoluble	Low	No	Yes	No	No	Yes	No	No	6.97

IV. CONCLUSION

In the presented work we have presented an artificial intelligence based technique for novel medicinal compound generation and tested it on recently released protease structure of SARS-CoV2. Finding medicinal compounds for Covid-19 was the initial trend during the onset of Covid-19 but such studies have its own limitation. This study outputs new formulation specifically for the input protease, followed by affectivity screenings and toxic analysis. According to our research, computational designing &screening is quite effective and this can supply possible repurposing drug candidates with results deliverable in just few days. It also recognizes a set of hot spot residues that make important contributions to protein-ligand binding, which can promote the rational design of novel protease inhibitors say SARS-CoV-2. This technique eliminate cut shorts the time invested in identifying compound by trial and error basis. AI system can formulate and check new compounds more than the speed of humans. This system can be latter deployed for finding cure for other major incurable diseases. The medicine compound generate by AI against corona can be rapidly synthesized in lab and in pharmaceutical companies for rapid building of vaccines. Sharing of our results will validate the results of other labs and pharmacy. We have not only be able to found 25 novel drug compounds against corona virus but also few novel anti-HIV, ant-tumor, anti-malarial, anti-arthritis drugs and handful of ant-inflammatory drugs in the process. We hope that this research will pave way for faster drug discovery, vaccine development in near future.

REFERENCES

[1] Liu, X.; Zhang, B.; Jin, Z.; Yang, H.; Rao, Z. Crystal structure of COVID-19 main protease in complex with an inhibitor N3. Protein DataBank, 2020, DOI: 10.2210/pdb6LU7/pdb.

[2] Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J. Z. H.; Hou, T. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. Chem. Rev. 2019, 119, 9478−9508.

[3] Wang, J. M.; Hou, T. J.; Xu, X. J. Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models. Curr. Comput.-Aided Drug Des. 2006, 2, 287−306.

[4] Wang, J.; Hou, T. Develop and test a solvent accessible surface area-based model in

conformational entropy calculations. J. Chem. Inf. Model. 2012, 52, 1199−1212.

[5] Law, V.; Knox, C.; Djoumbou, Y.; Jewison, T.; Guo, A. C.; Liu, Y.; Maciejewski, A.; Arndt, D.; Wilson, M.; Neveu, V.; Tang, A.; Gabriel, G.; Ly, C.; Adamjee, S.; Dame, Z. T.; Han, B.; Zhou, Y.; Wishart, D. S. DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res. 2014, 42, D1091−1097.

[6] Wang, J.; Ge, Y.; Xie, X. Q. Development and Testing of Druglike Screening Libraries. J. Chem. Inf. Model. 2019, 59, 53−65. (7)

[7] Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J. Med. Chem. 2004, 47, 1739−1749.

[8] Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. J. Am. Chem. Soc. 2001, 123, 5221−5230.

[9] Swanson, J. M.; Henchman, R. H.; McCammon, J. A. Revisiting free energy calculations: a theoretical connection to MM/PBSA and direct calculation of the association free energy. Biophys. J. 2004, 86, 67−74.

[10] Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. Validation and use of the MM-PBSA approach for drug discovery. J. Med. Chem. 2005, 48, 4040−4048.

[11] Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. J. Chem. Inf. Model. 2011, 51, 69−82.

[12] Xu, L.; Sun, H.; Li, Y.; Wang, J.; Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 3. The impact of force fields and ligand charge models. J. Phys. Chem. B 2013, 117, 8408−8421.

[13] Sun, H.; Li, Y.; Shen, M.; Tian, S.; Xu, L.; Pan, P.; Guan, Y.; Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring. Phys. Chem. Chem. Phys. 2014, 16, 22035−22045.

[14] Sun, H.; Li, Y.; Tian, S.; Xu, L.; Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. Phys. Chem. Chem. Phys. 2014, 16, 16719−16729.

[15] Chen, F.; Liu, H.; Sun, H.; Pan, P.; Li, Y.; Li, D.; Hou, T. Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein-protein binding free energies and re-rank binding poses generated by protein-protein docking. Phys. Chem. Chem. Phys. 2016, 18, 22129−22139.

[16] Karami, M.; Jalali, C.; Mirzaie, S. Combined virtual screening, MMPBSA, molecular docking and dynamics studies against deadly anthrax: An in silico effort to inhibit Bacillus anthracis nucleoside hydrolase. J. Theor. Biol. 2017, 420, 180−189.

[17] Chen, F.; Sun, H.; Wang, J.; Zhu, F.; Liu, H.; Wang, Z.; Lei, T.; Li, Y.; Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 8. Predicting binding free energies and poses of protein-RNA complexes. RNA 2018, 24, 1183−1194.

[18] Mishra, S. K.; Koca, J. Assessing the Performance of MM/PBSA, MM/GBSA, and QM-MM/GBSA Approaches on Protein/Carbohydrate Complexes: Effect of Implicit Solvent Models, QM Methods, and Entropic Contributions. J. Phys. Chem. B 2018, 122, 8113−8121.

[19] Sun, H.; Duan, L.; Chen, F.; Liu, H.; Wang, Z.; Pan, P.; Zhu, F.; Zhang, J. Z. H.; Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 7. Entropy effects on the performance of endpoint binding free energy calculation approaches. Phys. Chem. Chem. Phys. 2018, 20, 14450−14460.

[20] Wang, E.; Weng, G.; Sun, H.; Du, H.; Zhu, F.; Chen, F.; Wang, Z.; Hou, T. Assessing the performance of the MM/PBSA and MM/GBSA methods. 10. Impacts of enhanced sampling and variable dielectric model on protein-protein Interactions. Phys. Chem. Chem. Phys. 2019, 21, 18958−18969.

[21] Weng, G.; Wang, E.; Chen, F.; Sun, H.; Wang, Z.; Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 9. Prediction reliability of binding affinities and binding poses for

proteinpeptide complexes. Phys. Chem. Chem. Phys. 2019, 21, 10135−10145.

[22] Irwin, John J, and Brian K Shoichet. "ZINC--a free database of commercially available compounds for virtual screening." Journal of chemical information and modeling vol. 45,1 (2005): 177-82. doi:10.1021/ci049714.

[23] Kim, Sunghwan et al. "PubChem in 2021: new data content and improved web interfaces." Nucleic acids research vol. 49,D1 (2021): D1388-D1395.

[24] Wishart, David S et al. "DrugBank: a comprehensive resource for in silico drug discovery and exploration." Nucleic acids research vol. 34,Database issue (2006): D668-72. doi:10.1093/nar/gkj067

[25] Subramanian A, et al. A Next Generation Connectivity Map: L1000 Platform And The First 1,000,000 Profiles. Cell. 2017/12/1. 171(6):1437–1452.

[26] Rai, A., and J. Kannan. "SIMULATING DRUG-TARGET INTERACTION USING LARGE SCALE MOLECULAR DYNAMICS AND FUZZY-ART". Asian Journal of Pharmaceutical and Clinical Research, Vol. 10, no. 13, Apr. 2017, pp. 399-01, doi:10.22159/ajpcr.2017.v10s1.19973.

[27] Rai, A., and J. K. R. "PREDICTION OF CHRONIC BACTERIAL INFECTION BY IDENTIFICATION OF INTER CELLULAR RESPONSES OF GENETIC FUSION CENTERS". Asian Journal of Pharmaceutical and Clinical Research, Vol. 10, no. 13, Apr. 2017, pp. 417-9, doi:10.22159/ajpcr.2017.v10s1.19978.

[28] Soler, José M., et al. "The SIESTA method for ab initio order-N materials simulation." Journal of Physics: Condensed Matter 14.11 (2002): 2745.