Multi-reference Computational Method for De-novo Design, Optimization, and Repositioning of Pharmaceutical Compounds Illustrated by Identifying Multi-target SARS-CoV-2 Ligands

Vadim Alexandrov
Liquid Algo LLC https://orcid.org/0000-0002-0254-9799
Alexander Kirpich
Georgia State University https://orcid.org/0000-0001-5486-0338
Yuriy Gankin (yuriy.gankin@quantori.com)
Quantori https://orcid.org/0000-0003-0046-1037

Research Article

Keywords: conformers, multi-reference, poly-conformational, in silico, ligand-based, structure-based, SARS COV-2, COVID-19, fingerprints, cheminformatics, similarity, virtual library, computational framework, validation

Posted Date: September 3rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-153954/v2

Dear Journal of Cheminformatics Editorial Team,

We appreciate your consideration of our manuscript entitled "Multi-reference computational method for de-novo design, optimization, and repositioning of pharmaceutical compounds illustrated by identifying multi-target SARS-CoV-2 ligands".

In this paper, we have proposed the multi-reference optimization approach in various in-silico drug discovery settings and illustrated its application and usefulness. We have also performed the validation of our approach for SARS-CoV-2 compounds and the corresponding results have been included in our manuscript. According to the proposed approach, each molecule is represented as an ensemble of flexible conformers that would choose the best possible conformation for each presented target-binding opportunity that can be applied in multiple settings. The aims were: a) to present a universal search framework for potential candidate compounds based on the comparison of multiple similarities between compounds' conformers and b) to identify candidate compounds that are simultaneously similar to each of the selected known reference compounds. Application of this approach to SARS-CoV-2 produced several antiviral drug candidates that are designed to protect against SARS-CoV-2 by multiple mechanisms simultaneously.

Since this is a time sensitive matter, we appreciate your expedited consideration and review of our manuscript for publication in the Journal of Cheminformatics. Please do not hesitate to contact me with any additional inquiries regarding this submission.

On behalf of all authors,

Yuriy Gankin, Ph.D., MBA

Quantori. Custom Software Solutions.

625 Massachusetts Avenue

Cambridge, MA 02139

United States

E-mail: yuriy.gankin@quantori.com

Phone: (781) 771-1978

Multi-reference computational method for de-novo design, optimization, and repositioning of pharmaceutical compounds illustrated by identifying multi-target SARS-CoV-2 ligands

Vadim Alexandrov1, Alexander Kirpich2 and Yuriy Gankin*3

1 Liquid Algo LLC, 85 Thistle Lane, Hopewell Junction, NY 12533, United States of America

2 Department of Population Health Sciences, School of Public Health, Georgia State University, Atlanta, GA, United States of America

3 Quantori LLC, 625 Mass Ave, Cambridge, MA 02139, United States of America

*** Corresponding author**

E-mail: yuriy.gankin@quantori.com

ABSTRACT

In this work a novel computational multi-reference poly-conformational algorithm is presented for design, optimization, and repositioning of pharmaceutical compounds. The algorithm searches for candidates by comparing similarities between conformers of the same compound and identifies target compounds whose conformers are simultaneously "close" to the conformers for *each* of the compounds in a reference set. The reference compounds can have very different MoAs, which directly and *simultaneously* shapes the properties of the target candidate compounds.

The algorithm functionality has been validated *in silico* by scoring ChEMBL drugs against FDA-approved reference compounds which either had the highest predicted binding affinity to our chosen SARS-COV-2 targets or confirmed to be inhibiting such targets in-vivo. All our top scoring ChEMBL compounds also turned out to be either high-affinity ligands to the chosen targets (as confirmed separately in other studies) or showing significant efficacy in-vivo against those selected targets.

In addition to method validation *in silico* search for new compounds within two virtual libraries from the Enamine database is presented. The library's virtual compounds have been compared to the same set of reference drugs that we used for validation: Olaparib, Tadalafil, Ergotamine and Remdesivir. The large reference set of four potential SARS-CoV-2 compounds have been selected, since no drug has been identified to be 100% effective against the virus so far, possibly because each candidate drug was targeting only one particular MoA. The goal here was to introduce methodology for identifying potential candidate(s) that cover multiple MoA-s presented within a set of reference compounds.

KEYWORDS

conformers, multi-reference, poly-conformational, *in silico*, ligand-based, structure-based, SARS-COV-2, COVID-19, fingerprints, cheminformatics, similarity, virtual library, computational framework, validation.

INTRODUCTION

Conformers as independent molecular entities. In real life, most compound molecules exist in multiple conformations (shapes) based on the surrounding environmental conditions. In particular, each 3D shape of a molecule dictates its biological activity and enables the molecule to fit into the binding pockets of proteins. Often, distinctly different chemical compounds that have similar shapes (and similar charge distributions along the molecular surface) have a potential to bind as long as the ligand's partial charges are positioned in the binding pocket the same way (i.e., form the same hydrogen bonds). Therefore, it is beneficial to compare the shapes and surface distribution charges for target query and reference compounds on a conformer-by-conformer basis. If one of the conformers of the query molecule matches one of the conformers (especially bound-to-target) of the reference molecule, then there is a chance that the reference compound will also exhibit similar binding properties to the same target.

Alignment-free 3D-similarity scoring. OpenEye Scientific Software Inc. pioneered an algorithm and the corresponding tool ROCS [13] for comparing shapes of molecules by overlaying and measuring their molecular structures *in silico* and comparing differences between a query and target molecule. ROCS identifies potentially active compounds by comparing their shapes.

Moreover, the ROCS tool is competitive and often superior to structure-based approaches in virtual screening [14,15] both in terms of overall performance and consistency [16]. As a result, novel molecular scaffolds have been identified by using ROCS against various targets which have been considered very difficult to address computationally [17].

Challenges with overlaying. The process of molecular shapes overlaying remains computationally intensive and often is a bottleneck in the search process for similar molecules. This remains despite the recent so-called PAPER implementation of ROCS on GPU [18] and the development of FastROCS [19] for large (>1B) compound libraries. Recently, alternative methods for overlaying have been introduced as a substitute for the ROCS approach. The alternative overlaying is performed by comparing shape-based descriptors (a.k.a conformer-level 3D fingerprints). An example of such an approach is ElectroShape implemented in the ODDT package [20] and is based on the algorithm that incorporates shape, chirality, and electrostatics [21,22], and represents each conformer via a fixed-length vector of real-valued numbers. Similarly the E3FP package [23] also utilizes an alignment-invariant 3D representation of molecular conformers as a fixed-length binary vector for each conformer. These fingerprint-based approaches allow to calculate the similarity between two molecular shapes either as a Tanimoto distance (for binary fingerprints) or Euclidean distance (for real-valued fingerprints) computations. Such computations are orders of magnitude faster in comparison to alternative methods that require the actual alignment of the two compared conformers. Even though the calculation of a shape-based fingerprint for each conformer can be a rather computationally involved procedure, as soon as all conformers for the virtual library are fingerprinted and stored in a database, the similarity search for the query molecule in such a database is computationally quick. Therefore the computationally

efficient method proposed here is expected to be very useful for finding candidate drugs for multi-target disease indications, ligand-based drug design, and drug repurposing applications.

Method applications for SARS-CoV-2 treatment compounds. The set of SARS-CoV-2 treatment compounds have been used for *both* method validation since, there are compounds that have been confirmed to be effective [24] and for the search for new potential compounds based on the existing known set since no drug has been identified to be 100% effective against the virus, possibly because each candidate drug was targeting only one particular MoA[25]. The SARS-CoV-2 virus has been selected for method illustration because of the importance of the subject. The virus was introduced into the human population in the Chinese city of Wuhan in the Province of Hubei in December of 2019 [1–4]. Since then the epidemic of SARS-CoV-2 has rapidly spread Worldwide. The World Health Organization (WHO) has officially declared the SARS-CoV-2 pandemic in March 2020 just three months after its emergence [5]. The novel coronavirus received an official name SARS-CoV-2 and the virus pandemic was called COVID-19 [6]. The formal evaluation and comparison of SARS-CoV-2 drugs can be performed by studying the compound properties by treating patients and performing clinical trials [7–10,12] or by studying the properties of the corresponding compounds *in silico* [9,11] which is done in this work. For method validation we used public ChEMBL (version 28) database [26] to screen compounds against the most important viral targets, namely 3C-like protease (3CLpro, aka Main protease or Mpro), papain-like protease (PLpro) and RNA-dependent RNA polymerase (RdRp). These targets play a major role in the replication/transcription and host cell recognition and therefore, are vital for the viral reproduction and spread of infection. Because our method doesn't directly use target information but rather analyzes 3D shapes for a compound that was already predicted or experimentally found to be effective against a particular target (we call it a reference compound), one has to choose one (or

more) such compound(s) as a reference for each target. The focus for each of the above SARS-COV-2 targets (3CLpro, PLpro and RdRp) was on the reference compounds with the highest binding affinities from the recent in silico multi-target repurposing study.[24]

For the new compound search (virtual library screening) we used the same set of reference compounds as we used for the method validation.

METHODS

Representative conformer space and conformer-by-conformer comparison. The proposed computational algorithm extends the currently available methods [20–23] and introduces additional search flexibility via the use of the compound conformers. The proposal is to compare multiple possible shapes, adopted via varying environmental conditions, of the same molecule (i.e., conformers) rather than just a single shape that was used before. In particular, the suggested approach is based on the matching of ligand-ligand fingerprints without explicitly using target structure information unlike docking and molecular dynamics approaches that simulate physical binding of a ligand to the target. The supporting theory behind the method is based on the decision to treat conformers, which might have different binding characteristics and properties, as independent entities. In such an approach each conformer has the corresponding independent alignment-free 3D-similarity scoring using the known multi-references. All conformers were generated using the ETKDG algorithm implemented in RDkit [27]. Benchmarking studies have found ETKDG to be the best-performing freely available conformer generator up-to-date [28,29] providing diverse and chemically-meaningful conformers reproducing crystal conformations.

Unlike what the majority of computational methods had assumed a couple of decades or so ago (e.g. in the CoMFA method [58]), recent research indicates that the bioactive conformation is not necessarily the lowest-energy conformation in the presence of the receptor [59–61]. In particular, as long as an increase in energy for less favorable conformation is compensated by its binding to the target, i.e. the total ligand-target energy is lower than the sum of the energies for the non-bound target and ligand, the bound state is favored. The proposed method emphasizes and relies on this ligand's ability to use its potentially higher energy conformations depending on the target it attempts to bind. Note, however, that when a sufficiently large number of conformers is requested, ETKDG algorithm generates more conformers with lower energy than with higher energy [27,28], therefore when averaged over all conformers (and we generate 100 conformers per molecule), conformers with the lower energy will contribute more to the total overlap.

Actually, one of the things that distinguishes ligand-based 3D virtual screening methods from 2D methods is that one has to start worrying about how many conformers to include in the reference set. If the molecule is flexible, it can assume many shapes and pharmacophores. How to deal with this is one of the fundamental questions in ligand-based virtual screening (LBVS).

In a recent paper by Schrödinger team [30] Cappel *et al.* performed comprehensive benchmark analysis and found that the number of conformers needed for 3D LBVS is actually relatively low: 100 or less to achieve good performance. Thus, we used ETKDG to generate 100 conformers per molecule in this work.

The authors have called the approach MultiRef3D to emphasize that it is a fast, alignment-free multi-objective optimization protocol that maximizes the 3D overlap of a query molecule's conformational ensemble with conformational ensembles of multiple reference ligands. The

diagram of the proposed method is summarized in Fig. 1. The formal details of the approach are discussed further.

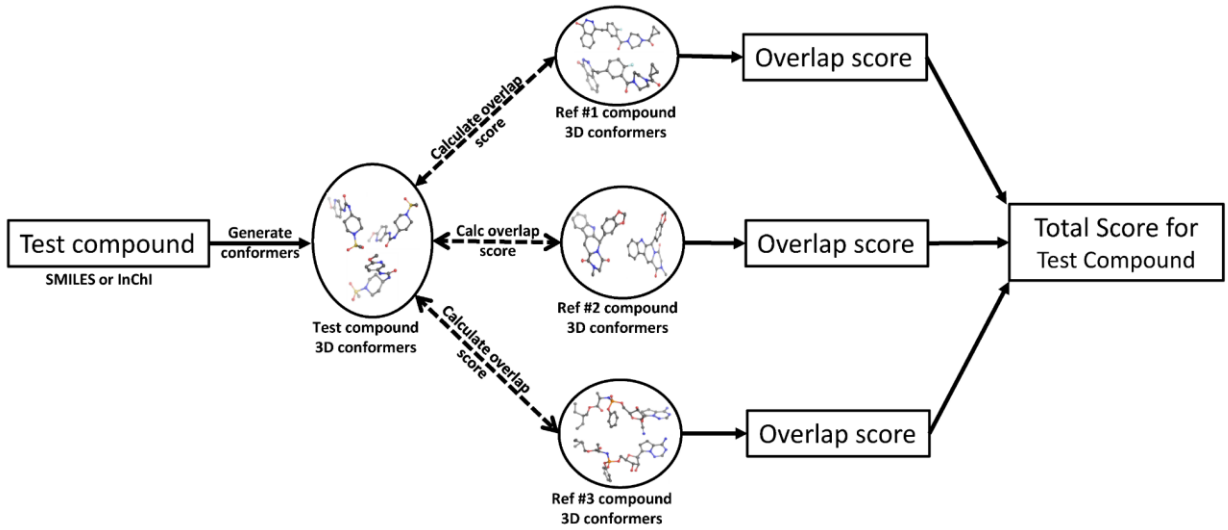

Fig. 1. MultiRef3D screening method diagram for multi-conformer and multi-reference screening procedure. For each test compound multiple conformers and the corresponding overlapping scores are computed. Later, the overlapping scores are summed into the total score for the selected test compound. Figure has been created in Microsoft PowerPoint 2016, pyMOL v2.5 (pymol.org) and RDKit v2021.03.01 software (rdkit.org).

Efficiency and a conformer scoring. In the algorithm, each conformation is treated as an independent entity and is characterized by a vector of features (fingerprint) which describes its 3D shape along with the distribution of electrostatic charge (both denoted further as electroshape) across its molecular surface. In this work we used 15-dimensional USRCAT fingerprints [31] which distil molecular shape into a rotation-invariant descriptor vector made up of 15 real numbers describing distance distribution among atoms, atomic partial charges and atom types. USRCAT

fingerprints were shown to significantly outperform just shape-based fingerprints in recent benchmark tests [31,32]. Since USRCAT fingerprints reflect both relative 3D positions for all atom types and molecular surface charges for each query molecule conformer as well as for all conformers of the reference compound they are very well-suited for alignment-free fast computation of conformer similarity. Each conformer is coded within the algorithm by a single fingerprint represented as a fixed-length vector of numbers which ensures computational efficiency. These fingerprints for each of the query and reference molecule conformers are individually scored by Euclidean distance serving as a similarity measure between two conformers. The Euclidean distance can be viewed as an extension of the Tanimoto similarity measure for non-binary fingerprints. The fingerprinting of individual conformers for alignment-free comparisons became popular in the past couple of years [23,33–35] so the proposed method is built on those.

Objective Function Optimization. The sum of the conformer-to-conformer similarity scores between the query and a reference compound are compared via an objective similarity function W_c for each reference compound c. The goal is to maximize the sum of those individual objective similarity functions across all reference compounds of interest $c=1,2,...,C$ where c is a summation index for the desired set of reference compounds:

$$W_{All} = \sum_{c=1}^{C} W_c = \sum_{c=1}^{C} \sum_{q=1}^{Q} \sum_{r=1}^{R} S_{q,r.}^{(c)} \qquad (1)$$

In formula (1) the summand $S_{q,r}^{(c)}$ is the similarity (overlap) of the query conformer q ($q=1,2,...,Q$) with the conformer r ($r=1,2,...,R$) for each reference compound c ($c=1,2,...,C$). For the real-valued fingerprints, the similarity summand between the pair of conformers of interest indexed by query index q and reference index r for compound c is calculated as:

$$S_{q,r}^{(c)} = 1 - (1/N)\sqrt{\sum_{n=1}^{N}(x_{q,n}^{(c)} - x_{r,n}^{(c)})^2} \quad (2)$$

where $x_{q,n}^{(c)}$ and $x_{r,n}^{(c)}$ are the corresponding normalized fingerprint vector coordinates for *n=1,2,...,N*. The length (the number of coordinates) of the fingerprint N is determined based on the problem-specific target-ligand interaction characteristics. Since the fingerprint coordinates $x_{q,n}^{(c)}$ and $x_{r,n}^{(c)}$ are normalized (i.e. have values between 0 and 1 for each coordinate *n*) the resulting overlap $S_{q,r}^{(c)}$ is maximized with the value equal to 1 when the fingerprints of both conformers are identical and can take the smallest value equal to 0 when all the fingerprint coordinates have a difference equal to 1 i.e. as different as possible at the normalized scale.

When the objective is to identify a novel compound for just a single active conformation (*r=1*) of one (*c=1*) reference compound (e.g. a reference ligand co-crystallized with one particular target) then all conformers for the query molecule are scored against only one active reference conformer. However, in the case when multiple reference compounds are bound to the same target (or sets of reference compounds bound to multiple targets), the total objective function comes into play. It is important to point out that the proposed method is not limited to the structure-based design situations: when several reference compounds are found to be active in a functional assay (and either the target(s) is unknown or the crystal structure of the target is not available) - the formula works just as well (as long as the ligand structure is known). The method becomes especially handy, when there is a great diversity among active reference compounds, whether the target structural information is known or not – the objective function will extract and sum up the similarities for all of the relevant parts of the fingerprinted conformer representations responsible for the observed activity.

The query compound can be evaluated against multiple reference compounds on a conformer-by-conformer basis. In such a case, the corresponding similarity scores are summed and constitute the multi-reference conformer-level objective function to maximize. This can be readily used in a typical ligand-based design setting. However, instead of just searching for a shape analog of one of the conformers of a reference compound, in the case of multiple references, the algorithm performs a search for such a compound in the virtual library whose conformers have overlapped with conformers of each of those reference compounds. The latter will increase the chances that the selected virtual compound binds the same way to the corresponding targets of each of the references (i.e. the selected compound is capable of forming conformations that resemble active conformations responsible for the MoA of each of the references).

Method validation for known targets. To validate the proposed methodology for the multi-target-specific conformer similarity the three following targets have been used: 3CLpro (Mpro), PLpro, and RdRp. The spike protein has not been included as the validation target since the pharmacological activity may not be correlated directly with the binding affinity to the interfacial site[24]. ChEMBL (version 28) public database[26] has been chosen as the universe for screening. The selected ChEMBL compounds were already marketed drugs for which at least one target is known. The corresponding ChEMBL extraction query is provided in the manuscript Supplement at GitHub (https://github.com/quantori/MultiRef3D). The screened set had a total of 2,604 compounds. The corresponding reference compounds for validation were selected from the recent multi-target *in silico* repurposing study[24] based on the highest binding affinities for each of the SARS-COV-2 three targets 3CLpro, PLpro and RdRp.

Compounds search based on reference compounds' conformers. One hundred conformers for each of the reference molecules were generated at the MMFF94 level of theory [40] and each

conformer was ODDT-fingerprinted [20] and saved in the MongoDB database [41]. The ODDT implementation [20] of ElectroShape fingerprints [22] has been selected to demonstrate the proposed approach because these fingerprints are considered to be state-of-the-art in ligand-based virtual screening experiments [32,42], and they are not limited to binary values.

Virtual libraries for screening. Virtual libraries (query compounds) for screening consisted of Enamine[41] focused "antiviral-like" set (3,995 compounds) and diverse Discovery Diversity Set (10,559 compounds) [44]. Molecules from each virtual library were simultaneously evaluated against several reference drugs with different MoA (3CLpro, PLpro and RdRp inhibition). A query molecule for which some of its conformers are similar in shape with conformers for all the reference drugs would receive a higher score. In this approach, multiple virtual compounds can be identified to have a good conformer overlap with conformers of the reference drugs.

RESULTS

Method Validation for SARS-CoV-2 Compounds. The highest affinity binder Olaparib (-9.2 kcal/mol) has been selected as a reference compound for 3CLpro, Tadalafil (-9.2 kcal/mol) for PLpro and Lumacaftor (-9.9 kcal/mol) for RdRp. However, when multi-target scoring against these three references has been performed, the top ten scoring compounds from ChEMBL *had no conformers* similar in 3D shape (Euclidean distance < 0.5) to Lumacaftor conformers. Therefore, the Lumacaftor reference has been replaced with the next best *in silico* RdRp binder[24]Ergotamine24 (-9.4 kcal/mol binding affinity to RdRp). The resulted scores produced by the proposed method are summarized in Tab. 1:

Tab. 1. Top ten scoring compounds showing simultaneous conformer similarity with the reference compounds Olaparib, Tadalafil, and Ergotamine.

Compound ID	Compound Name	TotalScore	Olaparib	Tadalafil	Ergotamine
CHEMBL779	Tadalafil	228.46	70.70	100.00	57.76
CHEMBL1737	Sildenafil citrate	225.15	81.30	58.34	85.50
CHEMBL521686	Olaparib	223.08	100.00	57.61	65.48
CHEMBL105442	Ci-1040	220.40	80.68	79.16	60.56
CHEMBL129857	As-602868	220.16	78.27	74.50	67.39
CHEMBL2037511	Epelsiban	219.86	81.58	70.28	68.01
CHEMBL565612	Sotrastaurin	219.13	79.93	69.36	69.83
CHEMBL1516474	Tegaserod maleate	217.83	80.22	76.56	61.05
CHEMBL1236682	Refametinib	217.78	76.01	81.57	60.20
CHEMBL1923502	Ulimorelin hydrochloride	217.56	76.29	74.79	66.47

Both Olaparib and Tadalafil had the highest scores which confirmed the previous finding [24] that these compounds are simultaneously good binders for *both* 3CLpro and PLpro. Our method has also picked up Sildenafil (more commonly known under the brand name Viagra) which just like Tadalafil (aka Cialis) is also known as a classical PDE5A inhibitor. Although those compounds are predominantly used in the treatment of male erectile dysfunction and pulmonary hypertension,

it was shown [45] that in the presence of SARS-COV-2 infection, PDE5 inhibitors prevent thromboembolism caused by inflammatory processes in COVID-19 patients via NO/cGMP pathway and are potent inhibitors of 3CLpro [46]

Ci-1040 and Refametinib are the other two hits from Tab. 1 and are potent MEK inhibitors with high 3D shape similarity to both Olaparib and Tadalafil. MEK inhibitors, including Olaparib[47] were recently demonstrated to reduce cellular expression of ACE2 while stimulating NK-mediated cytotoxicity and attenuating inflammatory cytokines during the severe stage of SARS-CoV-2 infection[48]. Ci-1040 was also previously shown to display a broad anti-influenza virus activity in vitro and to provide a prolonged treatment window compared to the standard of care in vivo, specifically in lung cells[49].

The other hit from Tab. 1 is Sotrastaurin which is a PKC inhibitor and has been experimentally shown to inhibit SARS-COV-2 replication in vivo [50]and has been found to be among the best 3CLpro binders during *in silico* ZINC database screening study [51] Yet another notable hit among the top ten selected compounds in Tab. 1 is As-602868: a potent IKK2 inhibitor. This class of compounds is currently preclinically tested for NF-kB mediated cytokine storm attenuation in severe COVID-19 patients [52].

The other top hit, Epelsiban, was originally developed as an oxytocin receptor agonist. However, it has been recently shown [54] that oxytocin plays a major role in activation of NF-kB-mediated pathways. Interestingly, recent research has revealed 50 that Remdesivir (in addition to being a potent RdRp inhibitor) is also reducing viral replication via NF-kB pathway. Therefore, this hit serves as an example of non-obvious 3D-shape-based drug repurposing idea generation linked to the relevant yet non-primary SARS-COV-2 inhibiting mechanisms of reference compounds.

285 Thus in our second validation experiment the RdRp reference compound Ergotamine has been
286 replaced with Remdesivir which, as we already mentioned, is a well-established RdRp inhibitor
287 and cytokine storm attenuator that works via NF-kB pathway. The resulted scores produced in the
288 second scoring setup are summarized in Tab. 2:

289 **Tab. 2**. Top ten scoring compounds showing simultaneous conformer similarity with the reference
290 compounds Olaparib, Tadalafil, and Remdesivir.

Compound ID	Compound Name	TotalScore	Olaparib	Tadalafil	Remdesivir
CHEMBL1694	Benazepril hydrochloride	180.82	66.67	64.26	49.89
CHEMBL515606	Cilazapril	180.61	64.56	61.56	54.50
CHEMBL495727	At-9283	179.03	68.17	56.15	54.71
CHEMBL2107495	Temafloxacin hydrochloride	178.94	67.15	55.78	56.01
CHEMBL1200779	Trovafloxacin mesylate	178.60	66.24	54.02	58.35
CHEMBL340978	Benoxaprofen	178.27	68.56	56.54	53.16
CHEMBL8	Ciprofloxacin	177.05	63.21	57.19	56.65
CHEMBL1200831	Spirapril hydrochloride	177.00	65.28	60.24	51.47
CHEMBL1201011	Quinapril hydrochloride	176.84	66.32	60.40	50.13
CHEMBL1168	Ramipril	176.54	65.32	63.26	47.96

291

For Olaparib, Tadalafil, and Remdesivir reference compounds half of the top ten hits (Benoxaprofen, Ciprofloxacin, Spirapril hydrochloride, Quinapril hydrochloride and Ramipril) turned out to be ACE inhibitors and coagulation modifiers acting via NF-kB related pathways [55,56]! In addition, all of them turned out to be also good binders of 3CLpro [57].

The other hits were Temafloxacin and Trovafloxacin, predicted to be potent 3CLpro ligands[55] and experimentally shown to inhibit virus replication [56,57], and anti-inflammatory drugs Benoxaprofen and Ciproflaxin predicted to target 3CLpro [58,59] as well.

Concluding the list is an interesting multi-target Aurora/JAK inhibitor hit: compound At-9283. JAK inhibitors, in general, have promising therapeutic potential for SARS-COV-2 treatment with their dual anti-inflammatory and anti-viral effects[60]. At-9283, however, has also been recently identified to reverse SARS-COV-2 transcriptomic signature [61] and due to its matching tipiracil's 3D pharmacophore scaffold also inhibits SARS-COV-2 Nsp15 endoribonuclease [62,63] and targets 3CLpro as well[64,65].

Virtual library screening for multi-target SARS-CoV-2 compounds. The results from the focused ("antiviral-like") and diverse ("Discovery Diversity Set") are summarized in Tab 1 and 2 respectively. . The algorithm visual summary is displayed in Fig. 1 for the W_{All} objective function. Tables 3 and 4 summarize the direct application results of the Enamine [41] focused "antiviral-like" and "Diverse Discovery Set" virtual library screening. The first two columns of the Tables contain query compound IDs and their computed overlap scores. The rows are sorted according to the total sum overlap score displayed in the second column.

Tab. 3. The top scoring compounds from the Enamine "antiviral-like" virtual library (the first column) are sorted by their total overlap score W_{All} (the second column). The values in the other columns correspond to the sums of the overlap scores of the conformers for the corresponding reference compounds.

Compound ID	**Wall**	**Olaparib**	**Tadalafil**	**Ergotamine**	**Remdesivir**
Z1693453146	254.11	71.68	56.30	76.76	49.38
Z434669842	248.84	66.18	55.33	69.92	57.41
Z1381427631	248.57	66.63	52.86	72.84	56.24
Z1381425049	247.97	66.09	53.03	71.98	56.88
Z1313285936	246.97	67.70	55.51	72.23	51.54
Z826278840	246.37	65.61	56.67	68.79	55.30
Z94559538	245.69	70.41	55.41	65.24	54.64
Z435640438	245.21	63.85	54.62	72.45	54.29
Z435642248	245.13	64.04	55.31	71.85	53.93
Z827564114	244.89	64.93	57.02	70.19	52.75

Tab. 4. The top scoring query compounds from the Enamine a "Diverse Discovery Set" virtual library (the first column) are sorted by their total overlap score W_{All} (the second column). The values in the other columns correspond to the sums of the overlap scores of the conformers for the corresponding reference compounds.

Compound ID	Wall	Olaparib	Tadalafil	Ergotamine	Remdesivir
Z1760146546	255.19	74.41	60.57	74.94	45.27
Z3077896041	254.26	67.67	57.38	75.53	53.67
Z2911083836	253.23	72.49	58.63	75.85	46.27
Z2446617864	252.49	73.18	59.52	75.06	44.72
Z1139281415	252.30	69.01	57.39	75.88	50.02
Z1354703942	251.79	77.22	61.63	80.63	32.32
Z2256366543	251.27	66.54	54.27	74.18	56.28
Z1139281396	250.99	68.40	57.90	74.43	50.27
Z1139280685	250.51	68.34	57.88	74.50	49.79
Z2959367287	250.32	69.54	55.40	72.80	52.58

For the visual illustration of the algorithm results two compounds with the highest scores from Tab. 3 and 4 have been presented in Fig. 2, panels A and B respectively. It is worth noting that these compounds are quite flexible molecules due to their amidebridge around which the ring substructures can rotate, which ensures the ability of those molecules to accommodate different targets.

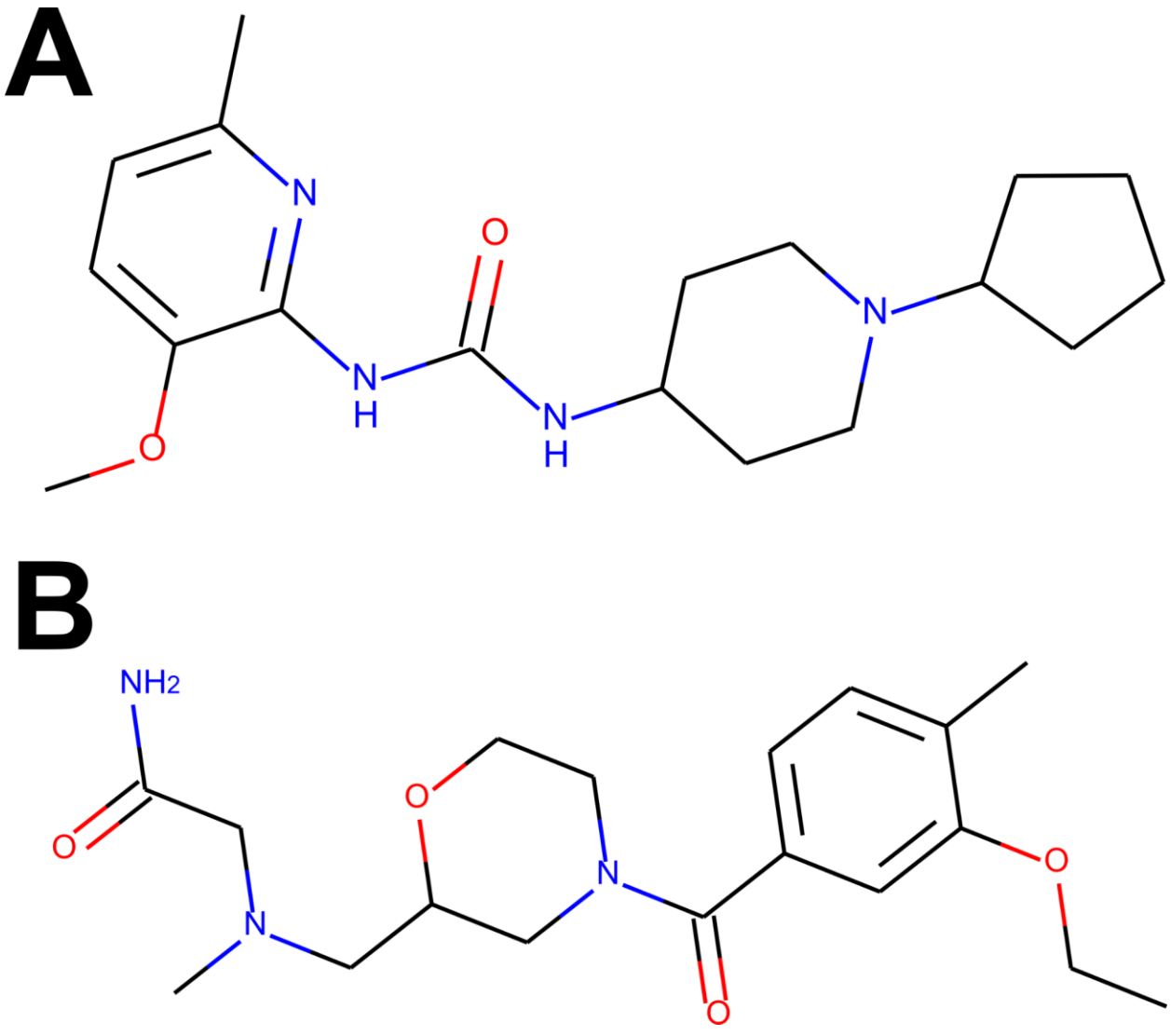

Fig. 2. The compounds presented in panels A and B are the top hits Z1693453146 (*Wall* = 254.11) and ZZ1760146546 (*Wall* = 255.19) from the non-overlapping "antiviral-like" and "Discovery Diversity" libraries respectively. One can immediately observe, however, that the compounds share a lot of similarity, in particular overall shape and amide bridge connecting heterocycles. The

bridge allows for 3D flexibility for the molecule to change conformation and bind to multiple targets. Figure has been created in Microsoft PowerPoint 2016, pyMOL v2.5 (pymol.org) and RDKit v2021.03.01 software (rdkit.org).

Fig. 3 demonstrates how the best-matching conformers of the top hit Z1693453146 spatially align with the active conformation for each reference drug. One can observe that the majority of hydrogen donors and acceptors from the top hit conformer and reference conformers are aligned very well mimicking the interaction patterns with each target. At least partial spatial alignment of atom types is expected from the top hit conformers since atom types as well as their relative 3D positions is the essence of the USRCAT fingerprints [31] .

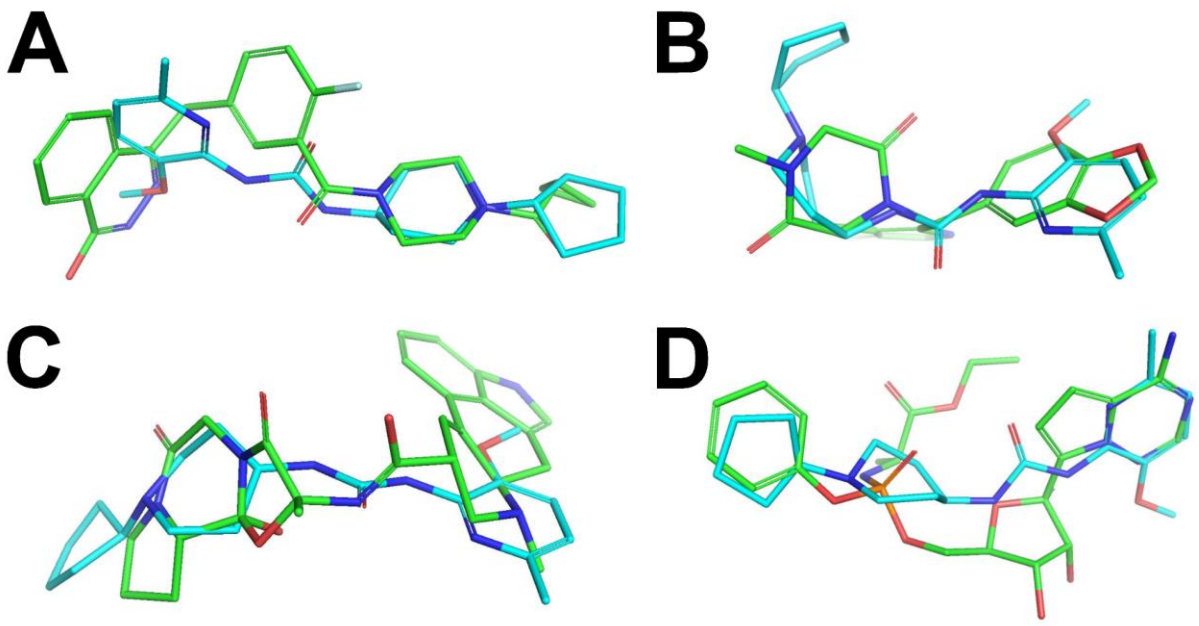

Fig. 3. Active conformers of the Reference compounds (Olaparib, Tadalafil, Ergotamine and Remdesivir on panels A, B, C and D respectively) aligned with the best matching conformers of the top hit Z1693453146 (Wall = 254.11). Carbon-carbon bonds for the reference compounds and

the tTop hHit are shown in green and cyan respectively (C-N and C-O bonds are conventionally shown in blue and red). Figure has been created in Microsoft PowerPoint 2016, pyMOL v2.5 (pymol.org) and RDKit v2021.03.01 software (rdkit.org).

DISCUSSION

Computation efficiency and availability of the method. The proposed method does not rely on laborious docking and molecular dynamics setup, especially in the multi-target case, where target preparation and choice of method i.e. direct docking to a fixed-coordinate target or Molecular Dynamics -based ensemble energy minimization are of utmost importance and require deep expertise. Fingerprint comparison is orders of magnitude faster and simpler (only requires simple structural information in the form of either isomeric SMILES or InChI). The entire setup is presented in our Supplement that can be universally used for any multi-target screening and optimization whenever reference compounds for each of the targets are available. Naturally, further hit refinement (ADMETox, PK/PD, etc) is necessary if the screened universe is not limited to the drugs with the well-known safety profiles.

Application for drug-repurposing. Depending on what is known about the indication or marketed drug of interest (targets, MoAs, other existing drugs for the same indication) the proposed methods (or a combination thereof) can be used to find other non-obvious molecules whose shape and the surface electrostatic charge is similar to that of the marketed drug. The methods can also be used to search for the cumulative similarity to conformers of the multiple drugs used to treat this disease indication.

In the proposed approach multiple conformers of the query ligand have been compared with conformers from *multiple* reference compounds whose therapeutic effect of interest is achieved via different mechanisms of bindings to different targets, e.g. by inhibiting major proteases 3CLpro and PLpro [63] and RNA-dependent RNA polymerase (RdRp) [6465,66]. An "ideal drug" would contain conformers that resemble (as many as possible) conformers of all of the reference drugs, thus increasing chances that the drug inhibits SARS-CoV-2 via multi-MoA routes and is more effective than each individual reference drug.

Note on applications for structure-based designs. When the crystal structure of the target protein is known and the reference ligand is co-crystallized in its active conformation (structure-based design), we can use this information about the reference compound and evaluate the query molecules against only one, the active (co-crystallized) reference ligand conformation ($r = r_{active}$) in formulas (1) and (2). Confirmation by direct docking for the fingerprint-matched queries can be used to confirm the match.

Our methodology emphasizes pursuit of candidate compounds that achieve therapeutic effect (e.g. stops SARS-CoV-2 proliferation) by multiple MoA routes. A successful candidate compound would contain conformers targeting the two major proteases 3CLpro and PLpro RdRpall at the same time by increasing chances that the compound would protect against SARS-CoV-2 much more effectively. Naturally, all successful candidates would need to be further screened and filtered for proper ADME-Tox and other drug-likeness properties. Binding to anti-targets, e.g. hERG, can be explicitly incorporated to this methodology by adding the corresponding terms (similarities to known hERG-binding ligands) to the overlap sum with a negative sign. Even though many computational methods exist to evaluate hERG in particular as well as other common tox liabilities, when an anti-target is very specific and less commonly known as "pure tox target"

(e.g. undesired binding to D2 receptor for many modern CNS drugs), the explicit inclusion of similarity score to such anti-target with a negative sign can greatly streamline the overall drug optimization process.

CONCLUSION

We have demonstrated and validated the usefulness of the multi-reference computationally efficient optimization approach in drug discovery screening and repurposing scenarios. The method represents each molecule as an ensemble of flexible conformers that would choose the best possible conformation for each presented target-binding opportunity. Application of this approach to SARS-CoV-2 produced several antiviral drug candidates that are designed to protect against SARS-CoV-2 by multiple mechanisms simultaneously.

LIST OF ABBREVIATIONS

ADME-Tox - Absorption, Distribution, Metabolism, Excretion and Toxicity

GPU - Graphics processing unit

CNS - Central nervous system

CoMFA - Comparative molecular field analysis

COVID-19 - Coronavirus Disease of 2019

CPU - Central processing unit

hERG - Human Ether-a-go-go-related Gene

MoA(s) - Mechanism of Action(s)

ODDT - Open Drug Discovery Toolkit

RNA - Ribonucleic acid

ROCS - Rapid overlay of chemical structures

SARS-CoV-2 - Severe acute respiratory syndrome coronavirus 2

WHO - World Health Organization

DECLARATIONS

Availability of data and materials

Code that has been used for analysis and for manuscript preparation can be found at Quantori public GitHub repository online [67]. Data (ligand structures) from REAL focused libraries can be downloaded from the Enamine Ltd. website [68].

Competing interests

The proposed method has been submitted for a patent. The patent application number is 63061790 at the United States Patent and Trademark Office and as of October 17, 2020, the patent is pending. The patent can be a source of financial income for authors Vadim Alexandrov (VA) and Yuriy Gankin (YG).

Funding

The author Vadim Alexandrov (VA), a founder of a consulting company Liquid Algo LLC in Hopewell Junction, New York, United States received no compensation for this work. The author Yuriy Gankin (YG) is employed by the commercial company Quantori in Cambridge, Massachusetts, United States. Alexander Kirpich (AK) is employed at the School of Public Health at a non-profit institution Georgia State University. Quantori provided support in the form of salary and relevant publication and patent fees for YG. YG received no compensation for this work. AK received no funding or any other financial support for this project.

Authors' contributions

Vadim Alexandrov (VA), Alexander Kirpich (AK), and Yuriy Gankin (YG) are the authors of the manuscript. VA and YG proposed the manuscript idea, obtained the data, implemented routine coding operations, and wrote the preliminary version of the manuscript. AK performed an additional literature review and wrote the final version of the manuscript. YG also provided the overall guidance for the project and participated in the manuscript preparation.

Acknowledgments

The authors would like to acknowledge Nika Tsutskiridze and Daviti Khatchilava for their assistance in extracting information from clinical trials and peer reviewed literature. The authors also want to acknowledge Alexander Proutsky and John Reynders for their suggestions and comments during the manuscript preparation.

Authors' information

VA holds a Ph.D. in Computational Chemistry from The University of Arizona and a Ph.D. in BioInformatics from Yale University. He has over 20 years of experience in computer-aided drug design and pioneered hit generation and lead optimization paradigm in high-throughput in-vivo phenotypic drug discovery.

AK is a biostatistician and an assistant professor in the Department of Population Health Sciences at the School of Public Health at Georgia State University. AK holds a Ph.D. in biostatistics from the University of Florida and has expertise in infectious disease statistical modeling (cholera, dengue, HIV, anthrax, and others), epidemiology, and bioinformatics. The interests and goals of AK are broad and target application of statistical methods to public health and biomedical research questions and policies and to address statistical challenges, such as missing data, asymptomatic infections, underreporting, and noise during data acquisition.

YG serves as Chief Scientific Officer at Quantori, a developer of intelligent IT and data science solutions for life science, pharma, and healthcare organizations. Yuriy's expertise is in analytical chemistry, data science, and chem- and bioinformatics. He holds a Ph.D. in Analytical Chemistry from Tufts University and an MBA from the Massachusetts Institute of Technology Sloan School of Management.

REFERENCES

1. Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727–733 (2020).
2. Lu, H., Stratton, C. W. & Tang, Y.-W. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. *J. Med. Virol.* **92**, 401–402 (2020).
3. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* vol. 395 497–506 (2020).
4. Chen, N. *et al.* Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet* vol. 395 507–513 (2020).
5. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-COVID-19---11-march-2020.
6. Naming the coronavirus disease (COVID-19) and the virus that causes it. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it.
7. Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S. C. & Di Napoli, R. Features, Evaluation, and Treatment of Coronavirus (COVID-19). in *StatPearls [Internet]* (StatPearls Publishing, 2020).
8. Sanders, J. M., Monogue, M. L., Jodlowski, T. Z. & Cutrell, J. B. Pharmacologic Treatments for Coronavirus Disease 2019 (COVID-19): A Review. *JAMA* (2020) doi:10.1001/jama.2020.6019.
9. Elmezayen, A. D., Al-Obaidi, A., Şahin, A. T. & Yelekçi, K. Drug repurposing for

coronavirus (COVID-19): screening of known drugs against coronavirus 3CL hydrolase and protease enzymes. *J. Biomol. Struct. Dyn.* 1–13 (2020).

10. Zhang, W. *et al.* The use of anti-inflammatory drugs in the treatment of people with severe coronavirus disease 2019 (COVID-19): The experience of clinical immunologists from China. *Clin. Immunol.* 108393 (2020).
11. Beck, B. R., Shin, B., Choi, Y., Park, S. & Kang, K. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. *Comput. Struct. Biotechnol. J.* **18**, 784–790 (2020).
12. Sargiacomo, C., Sotgia, F. & Lisanti, M. P. COVID-19 and chronological aging: senolytics and other anti-aging drugs for the treatment or prevention of corona virus infection? *Aging* **12**, 6511–6517 (2020).
13. Rocs, O. E. OpenEye Scientific Software, Inc., Santa Fe, NM, USA. (2008).
14. Hawkins, P. C. D., Skillman, A. G. & Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* **50**, 74–82 (2007).
15. Venhorst, J., Núñez, S., Terpstra, J. W. & Kruse, C. G. Assessment of scaffold hopping efficiency by use of molecular interaction fingerprints. *J. Med. Chem.* **51**, 3222–3229 (2008).
16. Sheridan, R. P., McGaughey, G. B. & Cornell, W. D. Multiple protein structures and multiple ligands: effects on the apparent goodness of virtual screening results. *J. Comput. Aided Mol. Des.* **22**, 257–265 (2008).
17. Rush, T. S., 3rd, Grant, J. A., Mosyak, L. & Nicholls, A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J. Med. Chem.* **48**, 1489–1495 (2005).

18. Haque, I. S. & Pande, V. S. PAPER--accelerating parallel evaluations of ROCS. *J. Comput. Chem.* **31**, 117–132 (2010).
19. FastROCS. OpenEye Scientific Software, Inc., Santa Fe, NM, USA. (2011).
20. Wójcikowski, M., Kukiełka, M., Stepniewska-Dziubinska, M. M. & Siedlecki, P. Development of a protein-ligand extended connectivity (PLEC) fingerprint and its application for binding affinity predictions. *Bioinformatics* **35**, 1334–1341 (2019).
21. Ballester, P. J. & Richards, W. G. Ultrafast shape recognition to search compound databases for similar molecular shapes. *J. Comput. Chem.* **28**, 1711–1723 (2007).
22. Armstrong, M. S. *et al.* ElectroShape: fast molecular similarity calculations incorporating shape, chirality and electrostatics. *J. Comput. Aided Mol. Des.* **24**, 789–801 (2010).
23. Axen, S. D. *et al.* A Simple Representation of Three-Dimensional Molecular Structure. *J. Med. Chem.* **60**, 7393–7409 (2017).
24. Murugan, N. A., Kumar, S., Jeyakanthan, J. & Srivastava, V. Searching for target-specific and multi-targeting organics for Covid-19 in the Drugbank database with a double scoring approach. *Sci. Rep.* **10**, 1–16 (2020).
25. Recent progress and challenges in drug development against COVID-19 coronavirus (SARS-CoV-2) - an update on the status. *Infect. Genet. Evol.* **83**, 104327 (2020).
26. Gaulton, A. *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40**, D1100–D1107 (2011).
27. Wang, S., Witek, J., Landrum, G. A. & Riniker, S. Improving Conformer Generation for Small Rings and Macrocycles Based on Distance Geometry and Experimental Torsional-Angle Preferences. *J. Chem. Inf. Model.* **60**, 2044–2058 (2020).
28. Friedrich, N.-O. *et al.* High-Quality Dataset of Protein-Bound Ligand Conformations and

549 Its Application to Benchmarking Conformer Ensemble Generators. *J. Chem. Inf. Model.* **57**,
550 529–539 (2017).
551 29. Friedrich, N.-O. *et al.* Benchmarking Commercial Conformer Ensemble Generators. *J.*
552 *Chem. Inf. Model.* **57**, 2719–2728 (2017).
553 30. Cappel, D., Dixon, S. L., Sherman, W. & Duan, J. Exploring conformational search
554 protocols for ligand-based virtual screening and 3-D QSAR modeling. *J. Comput. Aided*
555 *Mol. Des.* **29**, 165–182 (2014).
556 31. Schreyer, A. M. & Blundell, T. USRCAT: real-time ultrafast shape recognition with
557 pharmacophoric constraints. *J. Cheminform.* **4**, (2012).
558 32. Bonanno, E. & Ebejer, J.-P. Applying Machine Learning to Ultrafast Shape Recognition in
559 Ligand-Based Virtual Screening. *Frontiers in Pharmacology* vol. 10 (2020).
560 33. Wójcikowski, M., Zielenkiewicz, P. & Siedlecki, P. Open Drug Discovery Toolkit (ODDT):
561 a new open-source player in the drug discovery field. *J. Cheminform.* **7**, 26 (2015).
562 34. Gladysz, R. *et al.* Spectrophores as one-dimensional descriptors calculated from three-
563 dimensional atomic properties: applications ranging from scaffold hopping to multi-target
564 virtual screening. *J. Cheminform.* **10**, 9 (2018).
565 35. Wang, Y. *et al.* TF3P: Three-Dimensional Force Fields Fingerprint Learned by Deep
566 Capsular Network. *J. Chem. Inf. Model.* **60**, 2754–2765 (2020).
567 36. ClinicalTrials.gov. http://ClinicalTrials.gov.
568 37. Chen, X. & Geiger, J. D. Janus sword actions of chloroquine and hydroxychloroquine
569 against COVID-19. *Cell. Signal.* **73**, 109706 (2020).
570 38. Frediansyah, A., Nainu, F., Dhama, K., Mudatsir, M. & Harapan, H. Remdesivir and its
571 antiviral activity against COVID-19: A systematic review. *Clin Epidemiol Glob Health*

(2020) doi:10.1016/j.cegh.2020.07.011.

39. Agrawal, U., Raju, R. & Udwadia, Z. F. Favipiravir: A new and emerging antiviral option in COVID-19. *Armed Forces Med. J. India* **76**, 370–376 (2020).

40. Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry* vol. 17 490–519 (1996).

41. MongoDB Atlas Database. https://www.mongodb.com.

42. Cortés-Cabrera, A., Morris, G. M., Finn, P. W., Morreale, A. & Gago, F. Comparison of ultra-fast 2D and 3D ligand and target descriptors for side effect prediction and network analysis in polypharmacology. *British Journal of Pharmacology* vol. 170 557–567 (2013).

43. http://zinc15.docking.org/. http://zinc15.docking.org/.

44. REAL Database - Enamine. https://enamine.net/library-synthesis/real-compounds/real-database.

45. Shirvaliloo, M. Targeting the SARS-CoV-2 3CLpro and NO/cGMP/PDE5 pathway in COVID-19: a commentary on PDE5 inhibitors. *Future Cardiol.* (2021) doi:10.2217/fca-2020-0201.

46. Jin, Z. *et al.* Structure of M pro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289–293 (2020).

47. Vena, F. *et al.* MEK inhibition leads to BRCA2 downregulation and sensitization to DNA damaging agents in pancreas and ovarian cancer models. *Oncotarget* **9**, 11592–11603 (2018).

48. Zhou, L. *et al.* MEK inhibitors reduce cellular expression of ACE2, pERK, pRb while stimulating NK-mediated cytotoxicity and attenuating inflammatory cytokines relevant to SARS-CoV-2 infection. *Oncotarget* **11**, 4201 (2020).

595 49. The MEK-inhibitor CI-1040 displays a broad anti-influenza virus activity in vitro and
596 provides a prolonged treatment window compared to standard of care in vivo. *Antiviral Res.*
597 **142**, 178–184 (2017).
598 50. Liu, S. *et al.* Potential Antiviral Target for SARS-CoV-2: A Key Early Responsive Kinase
599 during Viral Entry. *CCS Chemistry* 559–568 (2021) doi:10.31635/ccschem.021.202000603.
600 51. Olubiyi, O. O., Olagunju, M., Keutmann, M., Loschwitz, J. & Strodel, B. High Throughput
601 Virtual Screening to Discover Inhibitors of the Main Protease of the Coronavirus SARS-
602 CoV-2. *Molecules* **25**, 3193 (2020).
603 52. Kircheis, R. *et al.* NF-κB Pathway as a Potential Target for Treatment of Critical Stage
604 COVID-19 Patients. *Front. Immunol.* **11**, (2020).
605 53. Hariharan, A., Hakeem, A. R., Radhakrishnan, S., Reddy, M. S. & Rela, M. The Role and
606 Therapeutic Potential of NF-kappa-B Pathway in Severe COVID-19 Patients.
607 *Inflammopharmacology* 1.
608 54. Oxytocin activates NF-κB-mediated inflammatory pathways in human gestational tissues.
609 *Mol. Cell. Endocrinol.* **403**, 64–77 (2015).
610 55. Hernández-Presa, M. *et al.* Angiotensin-converting enzyme inhibition prevents arterial
611 nuclear factor-kappa B activation, monocyte chemoattractant protein-1 expression, and
612 macrophage infiltration in a rabbit model of early accelerated atherosclerosis. *Circulation*
613 **95**, (1997).
614 56. Burzynski, L. C. *et al.* The Coagulation and Immune Systems Are Directly Linked through
615 the Activation of Interleukin-1α by Thrombin. *Immunity* vol. 50 1033–1042.e6 (2019).
616 57. Ísis Venturi Biembengut, T. de A. C. B. de S. Coagulation modifiers targeting SARS-CoV-
617 2 main protease Mpro for COVID-19 treatment: an in silico approach. *Mem. Inst. Oswaldo*

Cruz **115**, (2020).

58. Optibrium. Optibrium - StarDrop: Nova - A new generation of possibilities. https://www.optibrium.com/stardrop/stardrop-nova.php.

59. Gohda, K., Mori, I., Ohta, D. & Kikuchi, T. 10.1023/A:1008193217627. *Journal of Computer-Aided Molecular Design* vol. 14 265–275 (2000).

60. Mackerell, A. D., Jr. Empirical force fields for biological macromolecules: overview and issues. *J. Comput. Chem.* **25**, 1584–1604 (2004).

61. Hasegawa, K., Arakawa, M. & Funatsu, K. Rational choice of bioactive conformations through use of conformation analysis and 3-way partial least squares modeling. *Chemometrics Intellig. Lab. Syst.* **50**, 253–261 (2000).

62. Acharya, C., Coop, A., Polli, J. E. & Mackerell, A. D., Jr. Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. *Curr. Comput. Aided Drug Des.* **7**, 10–22 (2011).

63. Ullrich, S. & Nitsche, C. The SARS-CoV-2 main protease as drug target. *Bioorganic & Medicinal Chemistry Letters* vol. 30 127377 (2020).

64. Elfiky, A. A. SARS-CoV-2 RNA dependent RNA polymerase (RdRp) targeting: an perspective. *J. Biomol. Struct. Dyn.* 1–9 (2020).

65. Li, W. *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* vol. 426 450–454 (2003).

66. Vincent, M. J. *et al.* Chloroquine is a potent inhibitor of SARS coronavirus infection and spread. *Virol. J.* **2**, 69 (2005).

67. Website. *github.com - Quantori repository* https://github.com/quantori/MultiRef3D.

68. Targeted Libraries. https://enamine.net/hit-finding/focused-libraries/.

641