

Multi-reference poly-conformational computational methods for de-novo design, optimization, and repositioning of pharmaceutical compounds

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

Posted Date: December 4th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-120450/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

1 **Multi-reference poly-conformational computational methods for de-novo**
2 **design, optimization, and repositioning of pharmaceutical compounds.**

3

4 **Vadim Alexandrov¹, Alexander Kirpich² and Yuriy Gankin*³**

5

6 **1 Liquid Algo LLC, 85 Thistle Lane, Hopewell Junction, NY 12533, United States of America**

7 **2 School of Public Health, Georgia State University, Atlanta, Georgia, United States, Atlanta,**
8 **GA, United States of America**

9 **3 Quantori LLC, 625 Mass Ave, Cambridge, MA 02139, United States of America**

10 *** Corresponding author**

11 **E-mail: yuriy.gankin@quantori.com**

12

13

14

15

16

17

18

19

20 **ABSTRACT**

21 The COVID-19 epidemic, SARS-CoV-2, that began in December of 2019 has drastically altered
22 the aspects of daily life across the global society. Time-effective treatment of those infected has
23 since become a major goal with multiple treatment strategies having been designed to prevent the
24 progression of the disease into severe pneumonia. To date, no drug has been found to be 100%
25 effective against SARS-COV-2, possibly because each candidate drug was targeting only one
26 particular mechanism of action (MoA). Neither proposed up-to-date anti-SARS-COV-2 vaccine
27 are 100% effective. To contribute to the process of finding a more robust small-molecule solution,
28 utilizing several anti-SARS-COV-2 MoAs, a novel framework is presented; where the in silico
29 generated set of virtual library compounds is compared to six known reference drugs: Chloroquine,
30 Favipiravir, Remdesivir, JQ1, Apicidine, and Haloperidol which have been already used for
31 SARS-CoV-2 treatment. The aims were: a) to present a universal search framework for potential
32 candidate compounds based on the comparison of multiple similarities between compounds’
33 conformers and b) to identify candidate compounds that are simultaneously “close” to each of the
34 six known reference compounds that counteract SARS-CoV-2 via different mechanisms of action.

35

36 **KEYWORDS**

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

40

41

42 INTRODUCTION

43 **The necessity of the transparent algorithm.** The novel coronavirus SARS-CoV-2 was
44 introduced into the human population in the Chinese city of Wuhan in the Province of Hubei in
45 December of 2019 [1–4]. Since then the epidemic of SARS-CoV-2 has rapidly spread Worldwide.
46 The World Health Organization (WHO) has officially declared the SARS-CoV-2 pandemic in
47 March 2020 just three months after its emergence [5]. The novel coronavirus received an official
48 name SARS-CoV-2 and the virus pandemic was called COVID-19 [6]. As of November 2020,
49 there has been no virus-specific treatment for SARS-CoV-2 [7] but multiple drugs have been
50 proposed, such as remdesivir [8], talampicillin, lurasidone, rubitecan, loprazolam [9, 10],
51 chloroquine, and hydroxychloroquine [9, 10], atazanavir [11], azithromycin, quercetin,
52 chloroquine, rapamycin, and doxycycline [12]. The formal evaluation and comparison of drugs
53 can be performed by studying the compound properties by treating patients and performing clinical
54 trials [7–10, 12] or by studying the properties of the corresponding compounds in silico [9, 11].
55 The computational methods are preliminary and exploratory, but are less invasive than clinical
56 trials and allow the simultaneous study of large quantities of potential compounds. Since none of
57 the proposed small molecules were found to be 100%-effective, we present a new computational
58 method that combines the best features from the earlier individually imperfect small-molecule
59 treatments.

60 **Conformers as independent molecular entities.** In real life, most compound molecules exist in
61 multiple conformations (shapes) based on the surrounding environmental conditions. In particular,
62 each 3D shape of a molecule dictates its biological activity and enables the molecule to fit into the
63 binding pockets of proteins. Often, distinctly different chemical compounds that have similar
64 shapes (and similar charge distributions along the molecular surface) can bind as long as the

65 ligand's partial charges are positioned in the binding pocket the same way (i.e., form the same
66 hydrogen bonds). Therefore, it is beneficial to compare the shapes and surface distribution charges
67 for target query and reference compounds on a conformer-by-conformer basis. If one of the
68 conformers of the query molecule matches one of the conformers (especially bound-to-target) of
69 the reference molecule, then there is a chance that the reference compound will also exhibit similar
70 binding properties to the same target.

71 **Alignment-free 3D-similarity scoring.** OpenEye Scientific Software Inc. pioneered an algorithm
72 and the corresponding tool ROCS [13] for comparing shapes of molecules by overlaying and
73 measuring their molecular structures in silico and comparing differences between a query and
74 target molecule. ROCS identifies potentially active compounds by comparing their shapes.
75 Moreover, the ROCS tool is competitive and often superior to structure-based approaches in virtual
76 screening [14, 15] both in terms of overall performance and consistency [16]. As a result novel,
77 molecular scaffolds have been identified by using ROCS against various targets which have been
78 considered very difficult to address computationally [17].

79 **Challenges with overlaying.** The process of molecular shapes overlaying remains
80 computationally intensive and often is a bottleneck in the search process for similar molecules.
81 This remains despite the recent so-called PAPER implementation of ROCS on GPU [18] and the
82 development of FastROCS [19] for large (>1B) compound libraries. Recently, alternative methods
83 for overlaying have been introduced as a substitute for the ROCS approach. The alternative
84 overlaying is performed by comparing shape-based descriptors (a.k.a conformer-level 3D
85 fingerprints). An example of such an approach is ElectroShape implemented in the ODDT package
86 [20] and is based on the algorithm that incorporates shape, chirality, and electrostatics [21, 22],
87 and represents each conformer via a fixed-length vector of real-valued numbers. Similarly the

88 E3FP package [23] also utilizes an alignment-invariant 3D representation of molecular conformers
89 as a fixed-length binary vector for each conformer. These fingerprint-based approaches allow to
90 calculate the similarity between two molecular shapes either as a Tanimoto distance (for binary
91 fingerprints) or Euclidean distance (for real-valued fingerprints) computations. Such computations
92 are orders of magnitude faster in comparison to alternative methods that require the actual
93 alignment of the two compared conformers. Even though the calculation of a shape-based
94 fingerprint for each conformer can be a rather computationally involved procedure, as soon as all
95 conformers for the virtual library are fingerprinted and stored in a database, the similarity search
96 for the query molecule in such a database is computationally quick. Therefore the computationally
97 efficient method proposed here is expected to be very useful for finding candidate drugs for multi-
98 target disease indications, ligand-based drug design, and drug repurposing applications.

99

100 **METHODS**

101 **Conformer-by-conformer comparison.** The proposed computational algorithm extends the
102 currently available methods [20–23] and introduces additional search flexibility via the use of the
103 compound conformers. The proposal is to compare multiple possible shapes, adopted via varying
104 environmental conditions, of the same molecule (i.e., conformers) rather than just a single shape
105 that was used before. In particular, the suggested approach is based on the matching of ligand-
106 ligand fingerprints and goes beyond the known docking methods that utilize the simulated physical
107 binding of a ligand to the target. The supporting theory behind the method is based on the decision
108 to treat conformers, which might have different binding characteristics and properties, as
109 independent entities. In such an approach each conformer has the corresponding independent
110 alignment-free 3D-similarity scoring using the known multi-references. All conformers were

111 generated using the ETKDG algorithm implemented in RDkit [24]. Benchmarking studies have
112 found ETKDG to be the best-performing freely available conformer generator up-to-date [25, 26]
113 providing diverse and chemically-meaningful conformers reproducing crystal conformations.

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

117 **Efficiency and a conformer scoring.** In the algorithm, each conformation is treated as an
118 independent entity and is characterized by a vector of features (fingerprint) which describes its 3D
119 shape along with the distribution of electrostatic charge across its molecular surface. The
120 generated fingerprints reflect both 3D and surface charge (both denoted further as 3D+charge) for
121 each query molecule conformer as well as for all conformers of the reference compound. In this
122 case, each conformer is coded within the algorithm by a single fingerprint represented as a vector
123 of numbers which ensures computational efficiency. Those fingerprints for each of the query
124 molecule conformers are individually scored by Euclidean distance as a similarity measure
125 concerning each conformer of the reference compound. The Euclidean distance can be viewed as
126 an extension of the Tanimoto similarity measure for non-binary fingerprints. The fingerprinting of
127 individual conformers for alignment-free comparisons became popular in the past couple of years
128 [23, 27–29] so the proposed method is built on those.

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

134
$$W_{All} = \sum_{c=1}^C W_c = \sum_{c=1}^C \sum_{q=1}^Q \sum_{r=1}^R S_{q,r}^{(c)} \quad (1)$$

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

139
$$S_{q,r}^{(c)} = 1 - (1/N) \sqrt{\sum_{n=1}^N (x_{q,n}^{(c)} - x_{r,n}^{(c)})^2} \quad (2)$$

140 where $x_{q,n}^{(c)}$ and $x_{r,n}^{(c)}$ are the corresponding normalized fingerprint vector coordinates for
 141 $n=1,2,\dots,N$. The length (the number of coordinates) of the fingerprint N is determined based on
 142 the problem-specific target-ligand interaction characteristics. Since the fingerprint coordinates $x_{q,n}^{(c)}$
 143 and $x_{r,n}^{(c)}$ are normalized (i.e. have values between 0 and 1 for each coordinate n) the resulting
 144 overlap $S_{q,r}^{(c)}$ is maximized with the value equal to 1 when the fingerprints of both conformers are
 145 identical and can take the smallest value equal to 0 when all the fingerprint coordinates have a
 146 difference equal to 1 i.e. as different as possible at the normalized scale.

147 When the objective is to identify a novel compound for just a single active conformation ($r=1$) of
 148 one ($c=1$) reference compound (e.g. a reference ligand co-crystallized with one particular target)
 149 then all conformers for the query molecule are scored against only one active reference conformer.
 150 However, in the case when multiple reference compounds are bound to the same target (or sets of
 151 reference compounds bound to multiple targets), the total objective function comes into play. It is
 152 important to point out that the proposed method is not limited to the structure-based design
 153 situations: when several reference compounds are found to be active in a functional assay (and
 154 either the target(s) is unknown or the crystal structure of the target is not available) - the formula

155 works just as well (as long as the ligand structure is known). The method becomes especially
156 handy, when there is a great diversity among active reference compounds, whether the target
157 structural information is known or not – the objective function will extract and sum up the
158 similarities for all of the relevant parts of the fingerprinted conformer representations responsible
159 for the observed activity.

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

170 **Selection of the reference compounds and conformers.** In total six diverse compounds from
171 ClinicalTrials.gov [30] that are currently undergoing clinical trials for SARS-CoV-2 treatment
172 have been selected as reference compounds. These compounds are diverse in the sense that their
173 chemical structures differ substantially (i.e. Tanimoto distance < 0.7) from each other, and they all
174 have specificity for different targets, i.e. different Mechanism-of-Action (MoA). Three of these
175 compounds (Chloroquine [31], Remdesivir [32], and Favipiravir [33]) have recently demonstrated
176 significant efficacy against SARS-COV-2, whereas JQ1, Apicidin, and Haloperidol are already
177 marketed compounds well-known for their efficacy against other disease indications. One hundred

178 conformers for each of the reference molecules were generated at the MMFF94 level of theory
179 [34] and each conformer was ODDT-fingerprinted [20] and saved in the MongoDB database [35].
180 The ODDT implementation [20] of ElectroShape fingerprints [22] has been selected to
181 demonstrate the proposed approach because these fingerprints are considered to be state-of-the-art
182 in ligand-based virtual screening experiments [36, 37], and they are not limited to binary values.
183 Sometimes (e.g. at the beginning of a drug discovery program) a good virtual library might not
184 exist yet and fingerprinting the entire ZINC15 database [38] or Enamine REAL database [39] can
185 be computationally expensive (it may take several months to ODDT-fingerprint either of these
186 databases on a single CPU). Therefore in this work, two simplified approaches have been used to
187 illustrate the method. The first one is a focused virtual library approach that screened “only”
188 100,000 compounds from focused Enamine libraries, such as “anti-viral-like”. The second one is
189 an optimization based on a focused reference compound which is already known to be effective
190 for the target goals (e.g., start the optimization from a SARS-COV-2 reference compound, such as
191 chloroquine).

192 **Focused virtual library screening.** For the first approach, the virtual library (query compounds)
193 consisted of Enamine focused virtual sets (antiviral and other specialized libraries) from the
194 Enamine REAL database [39]. Molecules from this virtual library were simultaneously evaluated
195 against several antiviral reference drugs with different mechanisms of action (e.g. in the SARS-
196 COV-2 case the three major currently pursued MoA-s are: ACE2 binding, Mpro and RdRP
197 inhibition). A query molecule for which some of the conformers are similar in shape with
198 conformers for all the reference drugs would receive a higher score. In this approach, multiple
199 virtual compounds can be identified to have a good conformer overlap with the reference drugs.

200 The proposed in silico optimization algorithm from a focused reference compound (i.e.
201 chloroquine) is as follows:

- 202 1. Start in-silico synthesis (transformation) of any active compound in the training set by
203 applying chemical transformation rules from chemical reaction databases that ensure
204 tractability ("synthesizability") of each new "in-silico" compound obtained using these
205 rules (i.e each rule is a documented feasible transformation of a particular compound class
206 or compound substructure).
- 207 2. After each transformation, fingerprint all conformers from the obtained novel in-silico
208 compound and calculate the total overlap score with the reference conformers.
- 209 3. Based on the value of the total overlap score:
 - 210 a. reject novel compounds with low scores.
 - 211 b. keep transforming the top N (e.g. $N = 10$) highest scored compounds.
- 212 4. Stop the optimization procedure when the top score at the current iteration is no longer
213 higher than the top score in the previous iteration.

214

215 **RESULTS**

216 **Focused virtual library screening for SARS-CoV-2 compounds.** For the first approach, scoring
217 has been performed for the six compounds that are used for SARS-CoNV-2 treatment i.e.
218 Chloroquine, Remdesivir, Favipiravir, JQ1, Apicidine, and Haloperidol. The algorithm visual
219 summary is displayed in Fig. 1 for the W_{All} objective function. Table 1 summarizes the direct
220 application results of the focused antiviral Enamine virtual sub-library screening. The first two

221 columns of Table 1 contain query compounds from the Enamine REAL database [39] and their
 222 computed overlap scores. The rows are sorted according to the total sum overlap score displayed
 223 in the second column. The first two compared virtual compounds (Virt-cpd-001 through Virt-
 224 cpd-004) have the maximum sum of scores without “gaps”, i.e. none of the compounds have a
 225 score equal to zero indicating no overlap. This is different from compound Virt-cpd-014 whose
 226 conformers had no overlap with any of the conformers of Chloroquine, Remdesivir, and JQ1 but
 227 had a “good” overlap (in terms of the score) with conformers of Favipiravir, Apicidin, and
 228 Haloperidol.

229 **Table 1.** The query compounds from a virtual library (the first column) are sorted by their total
 230 overlap score W_{All} (the second column). The values in the other columns correspond to the sums
 231 of the overlap scores of the conformers for the corresponding reference compounds.

Virtual Compound	W_{All}	Choloroquine	Remdesivir	Favipiravir	JQ1	Apicidin	Haloperidol
Virt-cpd-001	354.37	58.25	58.30	59.14	59.38	59.34	59.94
Virt-cpd-002	283.28	48.73	24.64	36.07	50.53	61.31	62
Virt-cpd-003	282.95	48.1	25.8	37.26	49.85	60.61	61.34
Virt-cpd-004	269.63	37.87	23.46	46.09	39.12	61.06	62.03
Virt-cpd-005	241.03	29.27	0	59.51	29.87	60.58	61.78
Virt-cpd-006	209.92	34.42	34.61	34.93	34.94	35.35	35.67
Virt-cpd-007	182.34	21.04	22.94	38.17	21.41	38.08	40.7
Virt-cpd-008	173.60	45.31	20.82	0	46.62	21	39.85

Virt-cpd-009	172.77	32.23	32.1	0	45.28	29.81	33.36
Virt-cpd-010	168.55	32.29	11.72	16.28	33.49	31.1	43.67
Virt-cpd-011	165.05	0	18.2	31.89	20.03	42.59	52.34
Virt-cpd-012	155.42	39.85	0	11.76	34.63	34.39	34.8
Virt-cpd-013	152.10	39.57	0	0	39.44	26.97	46.12

232

233 When interpreting the results displayed in Table 1, the individual scores for each reference
234 compound should be taken into account along with the total score. For example, while compounds
235 Virt-cpd-012 and Virt-cpd-013 have very similar scores, conformers for Virt-cpd-014 have no
236 overlap with conformers of Remdesivir or Favipiravir, whereas conformers of Virt-cpd-012 have
237 no overlap only with conformers of Remdesivir. Since the objective is to find a compound whose
238 conformers have overlap with conformers for all of the reference compounds that have been
239 selected for the algorithm it is concluded that Virt-cpd-013 is substantially worse than Virt-cpd-
240 012. The other illustrative example is a total score comparison of Virt-cpd-005 vs Virt-cpd-006.
241 Even though the score for Virt-cpd-005 is higher than that for Virt-cpd-006, none of the conformers
242 of Virt-cpd-005 have any overlap with conformers of Remdesivir. The total score of Virt-cpd-005
243 is higher because the overlap of its conformers with those of Apicidin and Haloperidol is a lot
244 higher than the corresponding overlaps for Virt-cpd-006. Therefore based on Table 1. the
245 compounds Virt-cpd-006, Virt-cpd-007, Virt-cpd-008, Virt-cpd-011, Virt-cpd-012, and Virt-cpd-
246 013 are not recommended for the followup in-vivo tests since those compounds have no overlap
247 with any of the conformers for one of the reference compounds. A filter can be imposed for those
248 virtual compounds that have zero overlap with at least one of the references since the *multi-*

249 *objective* goal is to identify those virtual compounds that have the potential to exhibit *all* aspects
250 of the therapeutic intervention for each of the reference compounds. Therefore the compounds
251 with at least one such zero overlap are not recommended for the follow-up *in-vivo* testing
252 experiments.

253 For the visual illustration of the algorithm results two compounds with the highest scores from
254 Table 1. have been presented in Fig. 1. It is with noting that these compounds form very flexible
255 molecules due to their either sulphonyl or ether bridge around which the ring substructures can
256 rotate, which ensures the ability of those molecules to accommodate different targets.

257

258 Fig 1. The compounds presented in panels **A** and **B** from the antiviral Enamine virtual sub-library
259 collection that were found to maximize conformer overlap scores with the six reference
260 compounds. In addition to that sulphonyl bridge in panel **B** (circled in red) is a signature of the
261 classic antiviral compounds (e.g. well-known drug sulfapyridine), as well as the ether bond. The
262 bridge allows for 3D flexibility for the molecule to change conformation and bind to multiple
263 targets.

264 **In-silico optimization of a specific drug (chloroquine).** For the second approach chloroquine
265 drug has been selected for the illustration of the optimization from a focused single reference
266 compound. The selected chloroquine has been used as a starting point for the algorithm. For this
267 starting point, it has been discovered that the in-silico modification/optimization procedure quickly
268 converges to a series of novel compounds with high scores exceeding the highest score in the
269 original starting compound. The visual summary of the optimization results based on NOVA [40]
270 in-silico synthesis and selection for higher overall overlap are presented in Fig 2.

271

272 **Fig 2.** The original chloroquine is presented in panel **A**, while the chloroquine analogs optimized
273 in the modify-score-select algorithm are presented in panels **B** and **C**.

274

275 In Fig. 2 the compounds in panels **B** and **C** can be visually compared to the original chloroquine
276 structure in panel **A**. One can see that the two joined rings, the structural signature of chloroquine,
277 remained in the final optimized structures untouched by the optimization process (modifications
278 to that moiety would decrease the score and thus were rejected for the next iteration). This
279 approach can be viewed as building and scoring virtual library on-the-fly where the seed is chosen
280 to be one of the reference compounds. This can save a lot of compound generation scoring effort
281 (in contrast to scoring a comprehensive 140-billion conformation) library but might not hop too
282 far from the chemical neighborhood of the seed compound (the latter is not a drawback but rather
283 a desired feature during lead optimization effort around e.g. a patented chemotype or well-
284 researched chemical space in an already advanced lead optimization program).

285

286 **DISCUSSION**

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

293 **Special note on the use of non-lowest energy conformations and SARS-CoV-2.** Unlike what
294 the majority of computational methods had assumed a couple of decades or so ago (e.g. in the
295 CoMFA method [41]), recent research indicates that the bioactive conformation is not necessarily
296 the lowest-energy conformation in the absence of the receptor [42–44]. In particular, as long as an
297 increase in energy for less favorable conformation is compensated by its binding to the target, i.e.
298 the total ligand-target energy is lower than the sum of the energies for the non-bound target and
299 ligand, the bound state is favored. The proposed method emphasizes and relies on this ligand’s
300 ability to use its higher energy conformations depending on the target it attempts to bind.
301 Moreover, in the proposed method multiple conformers of the query ligand have been compared
302 with conformers from *multiple* reference compounds whose therapeutic effect of interest is
303 achieved via different mechanisms of bindings to different targets, e.g. by inhibiting Main protease
304 (Mpro) [45] and RNA-dependent RNA polymerase (RdRP) [46], while at the same time elevating
305 pH in lysosomes to arrest the intracellular proliferation of SARS-CoV-2 [47, 48]. An “ideal drug”
306 would contain conformers that resemble (some) conformers of all of the reference drugs, thus by
307 increasing chances that the drug inhibits SARS-CoV-2 via multi-MoA routes and is more effective
308 than each individual reference drug.

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

315 Our methodology emphasizes pursuit of candidate compounds that achieve therapeutic effect (e.g.
316 stops SARS-CoV-2 proliferation) by multiple MoA routes. A successful candidate compound
317 would contain conformers targeting M-protease, RdRP and endosome virus trafficking MoAs all
318 at the same time by increasing chances that the compound would protect against SARS-CoV-2
319 much more effectively. Naturally, all successful candidates would need to be further screened and
320 filtered for proper ADME-Tox and other drug-likeness properties. Binding to anti-targets, e.g.
321 hERG, can be explicitly incorporated to this methodology by adding the corresponding terms
322 (similarities to known hERG-binding ligands) to the overlap sum with a negative sign. Even
323 though many computational methods exist to evaluate hERG in particular as well as other common
324 tox liabilities, when an anti-target is very specific and less commonly known as “pure tox target”
325 (e.g. undesired binding to D2 receptor for many modern CNS drugs), the explicit inclusion of
326 similarity score to such anti-target with a negative sign can greatly streamline the overall drug
327 optimization process.

328

329 **CONCLUSION**

330 We have demonstrated the usefulness of the multi-reference optimization approach in various in-
331 silico drug discovery settings and illustrated its application for de-novo ligand-based design,
332 optimization, and repositioning of pharmaceutical compounds. The results allow the representation
333 of each molecule as an ensemble of flexible conformers that would choose the best possible
334 conformation for each presented target-binding opportunity that can be applied in multiple settings.
335 Application of this approach to SARS-CoV-2 produced several antiviral drug candidates that are
336 designed to protect against SARS-CoV-2 by multiple mechanisms simultaneously.

337

338

339

340

341 **LIST OF ABBREVIATIONS**

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

343 GPU - Graphics processing unit

344 CNS - Central nervous system

345 CoMFA - Comparative molecular field analysis

346 COVID-19 - Coronavirus Disease of 2019

347 CPU - Central processing unit

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

349 MoA(s) - Mechanism of Action(s)

350 ODDT - Open Drug Discovery Toolkit

351 RNA - Ribonucleic acid

352 ROCS - Rapid overlay of chemical structures

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

354 WHO - World Health Organization

355

356 **DECLARATIONS**

357 **Availability of data and materials**

358 Code that has been used for analysis and for manuscript preparation can be found at Quantori
359 public GitHub repository online [49]. Data (ligand structures) from REAL focused libraries can
360 be downloaded from the Enamine Ltd. website [50].

361 **Competing interests**

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

366 **Funding**

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

374 **Authors' contributions**

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

380 **Acknowledgments**

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

385 **Authors' information**

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

390 AK is a biostatistician and an assistant professor in the Department of Population Health Sciences
391 at the School of Public Health at Georgia State University. AK holds a Ph.D. in biostatistics from
392 the University of Florida and has expertise in infectious disease statistical modeling (cholera,
393 dengue, HIV, anthrax, and others), epidemiology, and bioinformatics. The interests and goals of
394 AK are broad and target application of statistical methods to public health and biomedical research

395 questions and policies and to address statistical challenges, such as missing data, asymptomatic
396 infections, underreporting, and noise during data acquisition.

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

402

403 **REFERENCES**

404

- 405 1. Zhu N, Zhang D, Wang W, et al (2020) A Novel Coronavirus from Patients with
406 Pneumonia in China, 2019. *N Engl J Med* 382:727–733
- 407 2. Lu H, Stratton CW, Tang Y-W (2020) Outbreak of pneumonia of unknown etiology in
408 Wuhan, China: The mystery and the miracle. *J Med Virol* 92:401–402
- 409 3. Huang C, Wang Y, Li X, et al (2020) Clinical features of patients infected with 2019 novel
410 coronavirus in Wuhan, China. *The Lancet* 395:497–506
- 411 4. Chen N, Zhou M, Dong X, et al (2020) Epidemiological and clinical characteristics of 99
412 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The*
413 *Lancet* 395:507–513
- 414 5. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March

- 415 2020. [https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-](https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-COVID-19---11-march-2020)
416 [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). Accessed 3 Sep 2020
- 417 6. Naming the coronavirus disease (COVID-19) and the virus that causes it.
418 [https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it)
419 [guidance/naming-the-coronavirus-disease-\(covid-2019\)-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).
420 Accessed 3 Sep 2020
- 421 7. Cascella M, Rajnik M, Cuomo A, et al (2020) Features, Evaluation, and Treatment of
422 Coronavirus (COVID-19). In: StatPearls [Internet]. StatPearls Publishing
- 423 8. Sanders JM, Monogue ML, Jodlowski TZ, Cutrell JB (2020) Pharmacologic Treatments for
424 Coronavirus Disease 2019 (COVID-19): A Review. JAMA.
425 <https://doi.org/10.1001/jama.2020.6019>
- 426 9. Elmezayen AD, Al-Obaidi A, Şahin AT, Yelekçi K (2020) Drug repurposing for
427 coronavirus (COVID-19): screening of known drugs against coronavirus 3CL hydrolase and
428 protease enzymes. J Biomol Struct Dyn 1–13
- 429 10. Zhang W, Zhao Y, Zhang F, et al (2020) The use of anti-inflammatory drugs in the
430 treatment of people with severe coronavirus disease 2019 (COVID-19): The experience of
431 clinical immunologists from China. Clin Immunol 108393
- 432 11. Beck BR, Shin B, Choi Y, et al (2020) Predicting commercially available antiviral drugs
433 that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep
434 learning model. Comput Struct Biotechnol J 18:784–790

- 435 12. Sargiacomo C, Sotgia F, Lisanti MP (2020) COVID-19 and chronological aging: senolytics
436 and other anti-aging drugs for the treatment or prevention of corona virus infection? *Aging*
437 12:6511–6517
- 438 13. Rocs OE (2008) OpenEye Scientific Software, Inc., Santa Fe, NM, USA
- 439 14. Hawkins PCD, Skillman AG, Nicholls A (2007) Comparison of shape-matching and
440 docking as virtual screening tools. *J Med Chem* 50:74–82
- 441 15. Venhorst J, Núñez S, Terpstra JW, Kruse CG (2008) Assessment of scaffold hopping
442 efficiency by use of molecular interaction fingerprints. *J Med Chem* 51:3222–3229
- 443 16. Sheridan RP, McGaughey GB, Cornell WD (2008) Multiple protein structures and multiple
444 ligands: effects on the apparent goodness of virtual screening results. *J Comput Aided Mol*
445 *Des* 22:257–265
- 446 17. Rush TS 3rd, Grant JA, Mosyak L, Nicholls A (2005) A shape-based 3-D scaffold hopping
447 method and its application to a bacterial protein-protein interaction. *J Med Chem* 48:1489–
448 1495
- 449 18. Haque IS, Pande VS (2010) PAPER--accelerating parallel evaluations of ROCS. *J Comput*
450 *Chem* 31:117–132
- 451 19. FastROCS (2011) OpenEye Scientific Software, Inc., Santa Fe, NM, USA
- 452 20. Wójcikowski M, Kukielka M, Stepniewska-Dziubinska MM, Siedlecki P (2019)
453 Development of a protein-ligand extended connectivity (PLEC) fingerprint and its
454 application for binding affinity predictions. *Bioinformatics* 35:1334–1341

- 455 21. Ballester PJ, Richards WG (2007) Ultrafast shape recognition to search compound
456 databases for similar molecular shapes. *J Comput Chem* 28:1711–1723
- 457 22. Armstrong MS, Morris GM, Finn PW, et al (2010) ElectroShape: fast molecular similarity
458 calculations incorporating shape, chirality and electrostatics. *J Comput Aided Mol Des*
459 24:789–801
- 460 23. Axen SD, Huang X-P, Cáceres EL, et al (2017) A Simple Representation of Three-
461 Dimensional Molecular Structure. *J Med Chem* 60:7393–7409
- 462 24. Wang S, Witek J, Landrum GA, Riniker S (2020) Improving Conformer Generation for
463 Small Rings and Macrocycles Based on Distance Geometry and Experimental Torsional-
464 Angle Preferences. *J Chem Inf Model* 60:2044–2058
- 465 25. Friedrich N-O, Meyder A, de Bruyn Kops C, et al (2017) High-Quality Dataset of Protein-
466 Bound Ligand Conformations and Its Application to Benchmarking Conformer Ensemble
467 Generators. *J Chem Inf Model* 57:529–539
- 468 26. Friedrich N-O, de Bruyn Kops C, Flachsenberg F, et al (2017) Benchmarking Commercial
469 Conformer Ensemble Generators. *J Chem Inf Model* 57:2719–2728
- 470 27. Wójcikowski M, Zielenkiewicz P, Siedlecki P (2015) Open Drug Discovery Toolkit
471 (ODDT): a new open-source player in the drug discovery field. *J Cheminform* 7:26
- 472 28. Gladysz R, Dos Santos FM, Langenaeker W, et al (2018) Spectrophores as one-dimensional
473 descriptors calculated from three-dimensional atomic properties: applications ranging from
474 scaffold hopping to multi-target virtual screening. *J Cheminform* 10:9

- 475 29. Wang Y, Hu J, Lai J, et al (2020) TF3P: Three-Dimensional Force Fields Fingerprint
476 Learned by Deep Capsular Network. *J Chem Inf Model* 60:2754–2765
- 477 30. ClinicalTrials.gov. <http://ClinicalTrials.gov>. Accessed 8 Sep 2020
- 478 31. Chen X, Geiger JD (2020) Janus sword actions of chloroquine and hydroxychloroquine
479 against COVID-19. *Cell Signal* 73:109706
- 480 32. Frediansyah A, Nainu F, Dhama K, et al (2020) Remdesivir and its antiviral activity against
481 COVID-19: A systematic review. *Clin Epidemiol Glob Health*.
482 <https://doi.org/10.1016/j.cegh.2020.07.011>
- 483 33. Agrawal U, Raju R, Udwardia ZF (2020) Favipiravir: A new and emerging antiviral option
484 in COVID-19. *Armed Forces Med J India* 76:370–376
- 485 34. Halgren TA (1996) Merck molecular force field. I. Basis, form, scope, parameterization,
486 and performance of MMFF94. *Journal of Computational Chemistry* 17:490–519
- 487 35. MongoDB Atlas Database. <https://www.mongodb.com>. Accessed 15 Oct 2020
- 488 36. Cortés-Cabrera A, Morris GM, Finn PW, et al (2013) Comparison of ultra-fast 2D and 3D
489 ligand and target descriptors for side effect prediction and network analysis in
490 polypharmacology. *British Journal of Pharmacology* 170:557–567
- 491 37. Bonanno E, Ebejer J-P (2020) Applying Machine Learning to Ultrafast Shape Recognition
492 in Ligand-Based Virtual Screening. *Frontiers in Pharmacology* 10
- 493 38. <http://zinc15.docking.org/>. <http://zinc15.docking.org/>. Accessed 15 Oct 2020

- 494 39. REAL Database - Enamine. <https://enamine.net/library-synthesis/real-compounds/real->
495 database. Accessed 15 Oct 2020
- 496 40. Optibrium Optibrium - StarDrop: Nova - A new generation of possibilities.
497 <https://www.optibrium.com/stardrop/stardrop-nova.php>. Accessed 15 Nov 2020
- 498 41. Gohda K, Mori I, Ohta D, Kikuchi T (2000) 10.1023/A:1008193217627. *Journal of*
499 *Computer-Aided Molecular Design* 14:265–275
- 500 42. Mackerell AD Jr (2004) Empirical force fields for biological macromolecules: overview
501 and issues. *J Comput Chem* 25:1584–1604
- 502 43. Hasegawa K, Arakawa M, Funatsu K (2000) Rational choice of bioactive conformations
503 through use of conformation analysis and 3-way partial least squares modeling.
504 *Chemometrics Intellig Lab Syst* 50:253–261
- 505 44. Acharya C, Coop A, Polli JE, Mackerell AD Jr (2011) Recent advances in ligand-based
506 drug design: relevance and utility of the conformationally sampled pharmacophore
507 approach. *Curr Comput Aided Drug Des* 7:10–22
- 508 45. Ullrich S, Nitsche C (2020) The SARS-CoV-2 main protease as drug target. *Bioorganic &*
509 *Medicinal Chemistry Letters* 30:127377
- 510 46. Elfiky AA (2020) SARS-CoV-2 RNA dependent RNA polymerase (RdRp) targeting: an
511 perspective. *J Biomol Struct Dyn* 1–9
- 512 47. Li W, Moore MJ, Vasilieva N, et al (2003) Angiotensin-converting enzyme 2 is a functional
513 receptor for the SARS coronavirus. *Nature* 426:450–454

- 514 48. Vincent MJ, Bergeron E, Benjannet S, et al (2005) Chloroquine is a potent inhibitor of
515 SARS coronavirus infection and spread. *Virology* 2:69
- 516 49. Website. In: github.com - Quantori repository. <https://github.com/quantori/MultiRef3D>.
517 Accessed 27 Nov 2020
- 518 50. Targeted Libraries. <https://enamine.net/hit-finding/focused-libraries/>. Accessed 27 Nov
519 2020
- 520

Figures

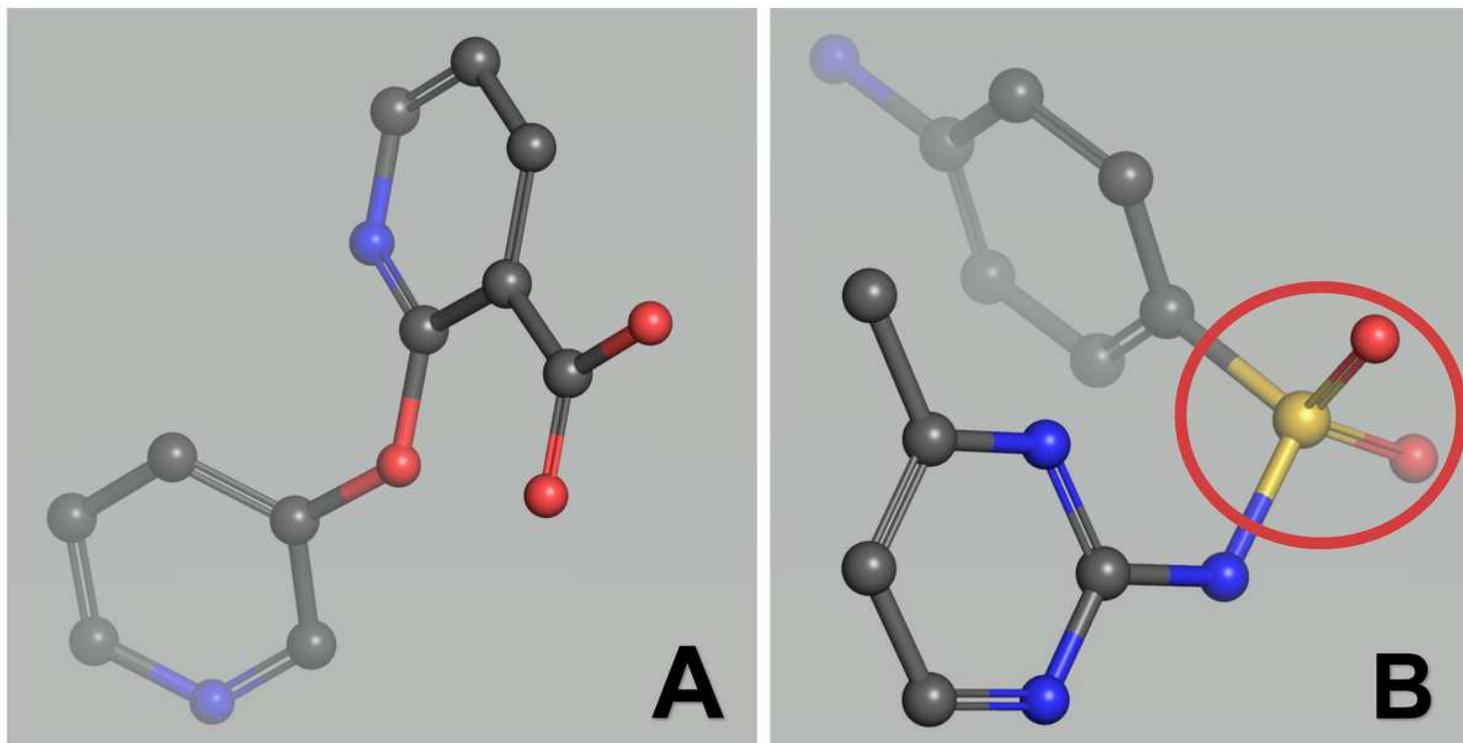


Figure 1

The compounds presented in panels A and B from the antiviral Enamine virtual sub-library collection that were found to maximize conformer overlap scores with the six reference compounds. In addition to that sulphonyl bridge in panel B (circled in red) is a signature of the classic antiviral compounds (e.g. well-known drug sulfapyridine), as well as the ether bond. The bridge allows for 3D flexibility for the molecule to change conformation and bind to multiple targets.

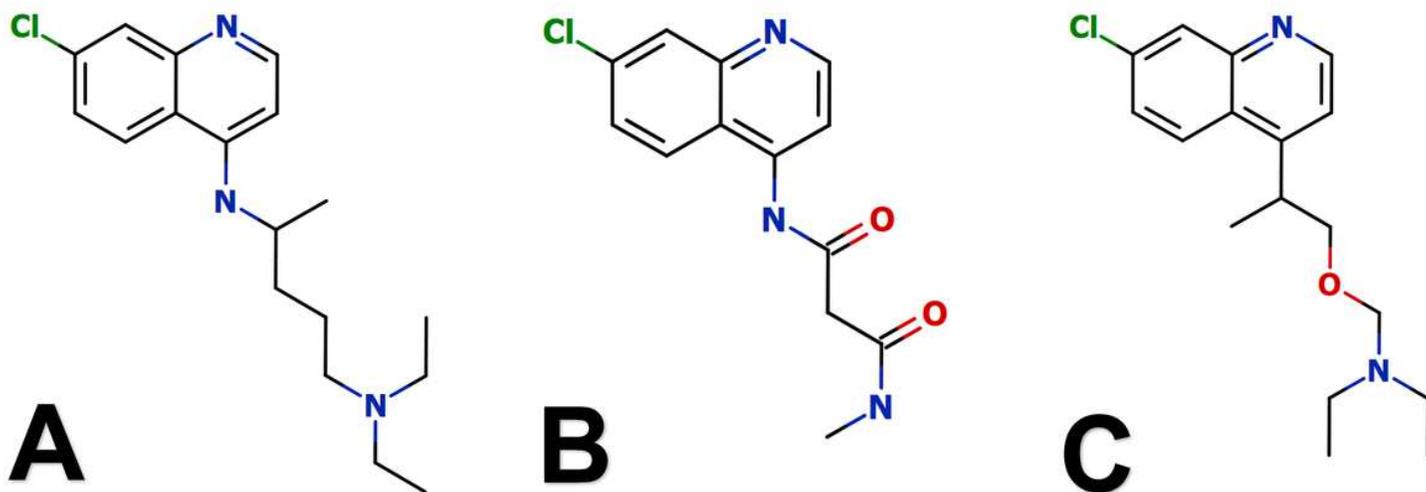


Figure 2

The original chloroquine is presented in panel A, while the chloroquine analogs optimized in the modify-score-select algorithm are presented in panels B and C.