 doi:10.20944/preprints202007.0558.v1

In silico study of SARS-CoV-2 Nucleocapsid Protein-Protein Interactions and Potential Candidates for their Stabilization

Julián F. Fernández[†] [‡] and Martín J Lavecchia[⋆]

[†]Universidad de Buenos Aires, Departamento de Química Orgánica, Facultad de Ciencias Exactas y Naturales, Ciudad Universitaria, Pabellón 2 (1428), Buenos Aires, Argentina.
[‡]CONICET-Universidad de Buenos Aires, Unidad de Microanálisis y Métodos Físicos en Química Orgánica (UMYMFOR), Buenos Aires, Argentina.
[⋆]CEQUINOR (UNLP—CONICET, CCT La Plata, associated with CIC PBA). Departamento de Química, Facultad de Ciencias Exactas, Universidad Nacional de la Plata. Blvd 120 1465 (1900) La Plata, Argentina.

July 22, 2020

 doi:10.20944/preprints202007.0558.v1

Abstract

The outbreak of the novel coronavirus SARS-CoV-2, the causative agent of COVID-19, has caused a global health crisis. Unfortunately, only a few treatments have proved to be effective, and their worldwide distribution remains as a challenge. Due to the urgency of the situation, drug repurposing remains as the fastest way to identify possible therapeutic options. Recent studies have shown that the stabilization of non-native Protein-Protein Interactions (PPIs) of the nucleocapsid protein of MERS coronavirus is a valid strategy to inhibit viral replication, but no study up to date has been done in SARS-CoV-2. In this work, a novel protocol for the discovery of PPIs stabilizers is presented and applied to SARS-CoV-2 N protein with a drug repurposing approach. This enabled us to identify that catechin, a structural motif present in widely distributed natural products, might be a privileged scaffold for this type of stabilization. Since many of the compounds presented in this work are generally considered nutraceuticals and have also been exhaustively studied, even though some of them contain PAINS substructures, could be good candidates for the SARS-CoV-2 nucleocapsid inhibition and be considered for further in vitro testing against COVID-19.

Contact: lavecchia@quimica.unlp.edu.ar
Supplementary information: Supplementary data available online.

1 Introduction

At the moment this article was written, more than 12 million positive cases of Coronavirus Disease 2019 (COVID-19) have been reported and the disease has already caused more than half a million deaths. Unfortunately, no vaccine has proven to be effective, and only few treatments have shown positive outcomes in their clinical trials, with their worldwide accesibility remaining as a challenge. Due to the urgency of the situation, drug repurposing is widely accepted as the fastest way to identify possible effective therapeutic options.

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), the causative agent of the disease, is a positive-sense single-stranded RNA virus belonging to the β genus of the Coronaviridae family. The SARS-CoV-2 virion consists of at least four (4) structural proteins: Spike (S) protein, membrane (M) protein, envelope (E) protein, and nucleocapsid (N) protein (Li *et al.*, 2020). Besides, some key non-structural proteins include: Papain like protease (PLpro) and Main protease (Mpro), which are responsible for the cleavage of viral polypeptides into functional units; and RNA-dependent RNA polymerase (RdRp), which is critical for viral proliferation (Ziebuhr *et al.*, 2000). As expected, these proteins have been identified as important drug targets (Dong *et al.*, 2020).

Among these, N-protein is a major structural protein that produces the ribonucleoprotein complex by the binding of viral RNA genome (Masters and Sturman, 1990). It plays an essential role in the regulation of viral RNA synthesis and also has functional importance in fundamental aspects of the CoV life cycle, such as the encapsidation and replication of virus genomes (Masters and Sturman, 1990). Moreover, its known that the disfunction or inhibition of this protein stops viral replication. The structure of these proteins is divided into three domains, namely an N-terminal

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0558.v1

RNA binding domain (NTD), a poorly structured central Ser/Arg (SR)-rich linker, and a C-terminal dimerization domain (CTD) (Chang *et al.*, 2013).

Recent studies showed that the stabilization of non-native Protein-Protein Interactions (PPIs) between these N terminal domains (NTD) in MERS coronavirus is a valid strategy to stop viral replication (Lin *et al.*, 2020). However, several structures of the SARS-CoV-2 N protein NTD have been reported since this research was published (Kang *et al.*, 2020, PDB 6VYO) and they reveal that there is only 58% similarity with their analogs in MERS. Therefore, this mechanism of action remains unexplored in SARS-CoV-2.

Modulation of PPIs by small molecules is an emerging and versatile strategy in drug development. The direct stabilization of these type of interactions is conceptually challenging since it requires the simultaneous targeting of more than one protein within the complex. Even though there are numerous examples of small molecules (especially natural products) that relay their biological activity on this type of interactions (Andrei *et al.*, 2017), very few studies have focused on modeling PPIs for drug design.

In this work, we identified three commercially available drugs that may be repurposed as potential non-native PPIs stabilizers between the NTDs of the SARS-CoV-2's nucleocapsid phosphoprotein. During the course of our investigation we also discovered that catechins, a specific type of flavonoids, are an excellent scaffold for this specific mechanism of action. Consequently, their interactions in comparison with previously reported biological activities were analyzed. We believe that the contributions made in this research may be an initial step for further research related to these compounds and also provides the basis for future small molecules aiming for PPIs stabilization development.

2 Methods

2.1 Interface Preparation

In order to obtain inducible homodimer interfaces between NTD N protein monomers, a crystal tetramer was selected (PDB ID 6VYO). This structure consists of four monomers stabilized with Zn^{2+} cations. The selected interfaces were covered by the union of chains A/B (**IF-1**) and chains B/C (**IF-2**), which are equivalent to the interface generated by chains C/D and D/A, respectively. Because there are missing residues (Uniprot P0DTC9), some of them close to the interface, the dimers were rebuilt by superimposing a monomeric experimental structure obtained by solution NMR (PDB ID 6YI3).

Furthermore, an additional interface was generated using the same methodology on 6KL5 dimer structure, a X-ray non-native dimer corresponding to the MERS-CoV N protein (Lin *et al.*, 2020). This dimer was called **IF-3** (see Figure 1).

Finally, these homodimers were subjected to a minimization, heating and NPT equilibration process in order to generate small cavities capable of harboring a stabilizing ligand.

2.2 Molecular Docking Studies

The compounds from the DrugBank 5.1.6 database (Wishart *et al.*, 2006) were screened against dimer interfaces. Molecular docking was performed to find and score binding poses on interfaces by means of FRED software, included in the OEDocking 3.4.0.2 suite, OpenEye (McGann, 2012). Parameters were kept to their de-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0558.v1

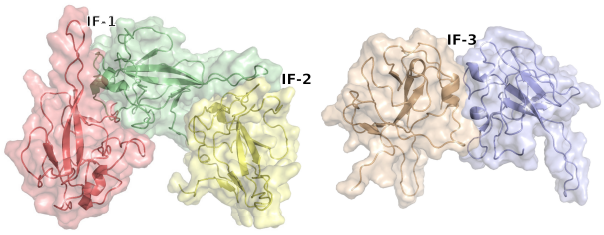

Figure 1: Selected homodimer interfaces to be stabilized: **IF-1** and **IF-2** built from 6VYO, and **IF-3** from 6KL5 structure. Image created using Pymol.

fault value. Based on the results, the Phenol-Explorer database (Neveu *et al.*, 2010) was included to be treated with this same docking protocol.

2.3 Molecular Dynamics (MD) Simulations

MD simulations were performed on complexes of the selected ligands and dimers. System topologies were built with the package AmberTools 19.0 (Case *et al.*, 2019) and all MD simulations were run using the NAMD 2.13 software (Phillips *et al.*, 2005). RMSD plots were depicted to determine the convergence and stability of simulations.

2.4 Interaction Energy Calculations

In order to quantify the dimer stabilization exerted by the ligand, two stabilization components were considered: ligand-dimer and monomer-monomer interactions. The sum of these two components was called Total Interaction. These energies were calculated using free energy decomposition analysis, which were performed using a pairwise energy decomposition scheme (idecomp option 3) with the MMPBSA.py module (Miller III *et al.*, 2012). This decomposition scheme also allowed the assignment of the ligand-dimer interaction percentage to each monomer, and thus ensure that there is a balanced distribution.

Additional information of this methodology and a general scheme of the protocol (see Figure S1) is detailed in the Supplementary information.

3 Results

3.1 Isolated Homodimers and their Interaction with Zn^{2+}

In order to compare the interface's behavior with and without a ligand, molecular dynamics simulations was first performed on the isolated dimers. These calculations revealed that monomers in **IF-1** stayed together during the MD trajectory. However, this was not observed in **IF-2** and **IF-3**. This phenomenon correlates with significant changes in monomer-monomer interaction energy and RMSD plots(see Tables S3 and S8).

With the intention of understanding the role played by Zn^{2+} in the crystal structure, it was incorporated to the calculated molecular dynamics and interaction energy calculations were performed. These results showed that this cation stabilizes both **IF-1** and **IF-2** dimers (see Table S8), which is an interesting discovery since this metal has aroused interest during the present pandemic due to its medicinal properties (Derwand and Scholz, 2020).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0558.v1

Table 1: Information on the selected compounds. An extended table is found in Supplementary information. Standard deviations are shown between parentheses. **Consensus* Docking position for compounds of DrugBank / DrugBank + Explorer-Phenol databases. n.i.: not included.

Name/Abbrev.	Type	*Consensus* Docking Position*	Interaction Energy **IF-1** (kcal/mol)	Interaction Energy **IF-2** (kcal/mol)	Interaction Energy **IF-3** (kcal/mol)
Masoprocol	Approved, Polyphenol	17 / 23	Dimer Separation	-76(5)	-185 (6)
Cianidanol	Approved, Catechin	38 / 10	-142 (6)	-84 (6)	-240 (8)
(+)-Gallocatechin Gallate / (+)GCG	Catechin	n.i. / 11	-134 (9)	-166 (8)	-129 (6)
(-)-Gallocatechin Gallate / (-)GCG	Catechin	n.i. / 55	-144 (7)	-133 (8)	Out of binding site
(-)-Epicatechine / (-)EC	Investigational, Catechin	18 / 27	-154 (8)	-150 (6)	-110 (6)
(+)-Gallocatechin / (+)GC	Catechin	n.i. / 8	-139 (7)	-120 (5)	-113 (6)
(-)-Epigallocatechin / (-)EGC	Catechin	3 / 7	-97 (6)	-103 (7)	-185 (6)
Ruxolitinib	Approved	26 / 18	-88 (5)	-114 (5)	Out of binding site
Quercetina	Flavonoid	27 / 59	-127 (6)	-88 (5)	Dimer Separation
(-)-Catechin Gallate / (-)CG	Catechin	n.i. / 32	-130 (6)	-162 (8)	-113(5)
(-)-Catechin / (-)C	Investigational, Catechin	n.i / 46	-127 (5)	-100 (7)	-125 (5)
(-)-Epigallocatechin Gallate / (-)EGCG	Investigational, Catechin	22 / 79	-118 (7)	-150 (6)	-116 (5)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0558.v1

3.2 Repurposed drugs as stabilizers

Once the interfaces were selected and the pockets generated, molecular docking calculations with both investigational and approved DrugBank's databases on all the interfaces were performed. At first, the high mobility of these systems seemed incompatible with the rigid nature of standard docking calculations. To address this situation, the compounds were ranked according to a *consensus* docking score, constructed from the relative docking position of each compound in each interface. Another important issue at this point each system was being considered as equally probable to be induced, which was not necessarily true. Since RNA assembly is the main function of this protein (and in consequence, its inhibition represents a possible mechanism of action), it was considered that a good strategy to solve this problem would be to prioritize those ligands capable to interact with the residues involved in this process. With this in mind, a fourth system focused on those residues was added to the previous *consensus* docking score ranking (Table S1, Column "Score RNA and Pos. RNA").

Rational manual selection of the best candidates was made. For this task, the best approved ligands in this ranking were given priority over the investigational ones, and those compounds that had a top score in any of the interfaces were also considered, even if their *consensus* ranking was not the best. With this strategy, 13 compounds were selected (Table S1) for molecular dynamics.

Stabilization was accomplished if ligand, interface-forming residues and protein RMSD reached a plateau, and also if the compound showed balanced interactions with residues from both monomers, characterized with MM-GBSA energy decomposition (see Table S9). These simulations showed that the compounds that were selected because of their high docking score in only one of the interfaces were not able to stabilize it, and therefore were not good candidates for further studies. However, most molecules selected from the *consensus* docking score ranking were able to stabilize two or more interfaces correctly, and are shown in Table 1. From this list, Quercetin and Ruxolitinib are already being tested in clinical trials against COVID-19, with results still pending (ClinicalTrials.gov ID NCT04377789 and NCT04348071).

3.3 Catechins as Promising Candidates

Our previous docking results showed that most of the molecules with the lowest score were mostly polyphenols, in particular flavonoids. Moreover, molecular dynamics simulations and energy calculations showed that all of these molecules stabilized at least two interfaces. This is particularly interesting since polyphenolic compounds are phytochemicals widely spread throughout plants and fruits, and are known to have medicinal and chemopreventive activities, for example they are good natural antioxidants. Interestingly, some previous studies have reported that some polyphenols inhibit nucleocapsid phosphoprotein's function in SARS-CoV (Roh, 2012), which has a 90.5% of identity with SARS-CoV, according to BLAST. In this work, Roh compares the binding affinity of an engineered RNA oligonucleotide with several polyphenols that block its interaction with the protein, and therefore inhibit its functionality. These findings motivated the addition of a polyphenol-specific database to the present study.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0558.v1

Molecular docking calculations of these compounds showed that, among the 378 polyphenolic molecules in the new database, flavonoids had the lowest *consensus* scores. Particularly, when these results were added to the previous ones obtained in this work, it was observed that 8 of the best 50 docked molecules were catechins (Table S1). Moreover a good differentiation between the compounds tested as active and inactive in Roh's work was found, particularly considering the **IF-2** docking score (Table S2). These results motivated the addition of (+)-Gallocatechin Gallate, (+)-Gallocatechin, (-)-Catechin Gallate for further molecular dynamics simulations. Also (-)-Catechin and (-)-Gallocatechin Gallate were taken into account for comparison, since they had been tested in Roh's work. Finally, despite not being a catechin itself, Gallic acid 3-O-gallate was also taken into account because it is a synthetic precursor of many of the molecules selected, and it had the second best *consensus* docking score of all.

Molecular dynamics simulations resulted in line with our previous calculations, since all catechins tested stabilized at least two interfaces (see Table S9). However, this was not the case for Gallic acid 3-O-gallate, which left the binding site in **IF-2** and **IF-3**. This result highlights the importance of the common catechin heterocyclic motif for this type of stabilization.

3.4 Energy Calculations

The final step in this protocol consists of interaction energy estimation. As mentioned in the Methods section, both ligand-dimer and monomer-monomer interactions were calculated, and their global contributions (Table 1, column "Interaction Energy") were considered.

The simulations performed up to this point showed that several compounds effectively stabilize one or more of the studied interfaces. At first, it was thought that the lack of experimental data made it almost impossible to establish a cutline and select the most promising compounds from this list. However, an excellent correlation was found between previously known active/inactive compounds and the calculated interaction energies in **IF-1** and **IF-2**. This was not the case for **IF-3**, however this result makes sense since this dimer does not include the residues that are known to interact with RNA, which were specifically tested in the biological assay.

Energy decomposition is presented in Figure 2. In these graphics, the most important ligand-residue interactions for each interface can be easily observed. An strategy for their analysis in **IF-1** and **IF-2** is to focus on those residues that are known to interact with RNA (Kang *et al.*, 2020).

Focusing on these residues in **IF-1**, it can be seen that His59, Arg92, Ile94, Ser105 and Tyr172 from chain A, and Arg149 from chain B, and are part of this binding site (Figure 2). Even though few patterns can be observed, it can be observed that (-)EGCG, (-)CG, (-)EC and (-)C showed strong interactions with many of these key residues, whereas the other ligands did not. Finally, Glu174 (chain B) showed a remarkable interaction with most of the compounds, and therefore proved to be an important residue for this dimer stabilization.

On the other hand, the number of RNA-binding residues in **IF-2** is slightly higher. In its pocket, Thr57, His59, Arg92, Ile94, Ser105, Arg107, Tyr172 from chain B showed significant interaction with many of the compounds, especially with those that had a better overall interaction energy. Moreover, when compar-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0558.v1

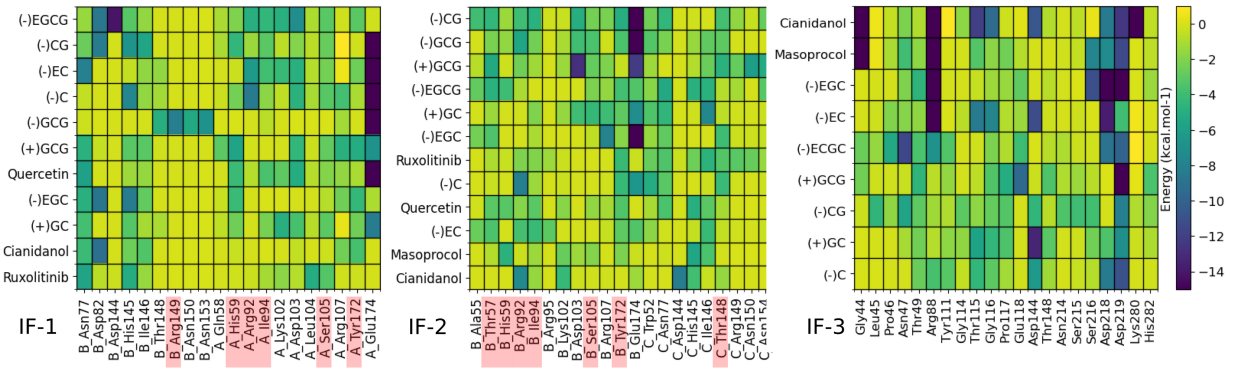

Figure 2: Per-residue decomposition analyses for selected compound of ligand-dimer energy component. The rows are ordered according to the total ligand-dimer contribution, decreasing downwards. In addition, the residues that interact with RNA, according to Kang *et al.*, 2020, are highlighted. Compounds full name are showed in Table 1.

ing these results with the activities measured in Roh's work, it can be observed that the active compounds (-)GCG and (-)CG interact significantly better with these residues in comparison with the inactive molecules (-)C and Quercetin, which could explain the difference between their biological activities. Finally, Glu174 shows, once again, a strong interaction with those ligands with better performance, highlighting its crucial role in both interfaces stabilization.

Lastly, **IF-3** proved to be a completely different interface than the previous two. No RNA-binding residues are found in its pocket, but other residues turned out to be important for its stabilization. These are: Gly44, Arg88, Asp144 from chain A and Asp81, Asp82 from chain B. In addition, some of the best ligands in this system, such as Cianidanol and Masoprocol, showed very weak ligand-dimer interaction in the previous two interfaces, suggesting that they may be able to inhibit this proteins function with a different mechanism than the rest.

4 Conclusions

The protocol developed in the present work enabled the identification of good candidates for PPI stabilization, which may also work as non-tested monomeric inhibitors (Table 1 and Figure S2). Therefore, besides the previous known active compounds (-)GCG and (-)CG, the new results suggests that Cianidanol, (-)EC, Masoprocol, (-)EGC, (-)EGCG, (+)GCG and (+)GC, even though some of them contain PAINS substructures, could be good candidates for the SARS-CoV-2 nucleocapsid inhibition. Considering the global health crisis created by SARS-CoV-2 has arisen and the present absence of any globally distributed treatment, these findings may promote further in vitro testing of these compounds.

Acknowledgements

JFF thanks the University of Buenos Aires for a doctoral fellowship. We gratefully acknowledge

the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research and OpenEye for providing an academic license to use their suite. Also, we thank Dr. J. A. Palermo for his invaluable help during this research.

Funding

This work was supported by CONICET and UNLP. MJL is member of the Carrera del Investigador CONICET.

References

Andrei, S. A. *et al.* (2017). Stabilization of protein-protein interactions in drug discovery. *Expert Opinion on Drug Discovery*, **12**(9), 925–940.

Case, D. *et al.* (2019). Ambertools19.

Chang, C.-k. *et al.* (2013). Transient oligomerization of the sars-cov n protein–implication for virus ribonucleoprotein packaging. *PloS One*, **8**(5), e65045.

Derwand, R. and Scholz, M. (2020). Does zinc supplementation enhance the clinical efficacy of chloroquine/hydroxychloroquine to win todays battle against covid-19? *Medical Hypotheses*, page 109815.

Dong, S. *et al.* (2020). A guideline for homology modeling of the proteins from newly discovered betacoronavirus, (2019-ncov). *J Med Virol*.

Graham, H. N. (1992). Green tea composition, consumption, and polyphenol chemistry. *Preventive Medicine*, **21**(3), 334–350.

Kang, S. *et al.* (2020). Crystal structure of sars-cov-2 nucleocapsid protein rna binding domain reveals potential unique drug targeting sites. *Acta Pharm Sin B*.

Li, X. *et al.* (2020). Molecular immune pathogenesis and diagnosis of covid-19. *JPA*.

Lin, S.-M. *et al.* (2020). Structure-based stabilization of non-native protein–protein interactions of coronavirus nucleocapsid proteins in antiviral drug design. *J Med Chem*, **63**(6), 3131–3141.

Masters, P. S. and Sturman, L. S. (1990). Background paper functions of the coronavirus nucleocapsid protein. In *Coronaviruses and their Diseases*, pages 235–238. Springer.

McGann, M. (2012). Fred and hybrid docking performance on standardized datasets. *J Comput Aid Mol Des*, **26**(8), 897–906.

Miller III, B. R. *et al.* (2012). Mmpbsa. py: an efficient program for end-state free energy calculations. *J Chem Theory Comput*, **8**(9), 3314–3321.

Neveu, V. *et al.* (2010). Phenol-explorer: an online comprehensive database on polyphenol contents in foods. *Database*, **2010**.

Phillips, J. C. *et al.* (2005). Scalable molecular dynamics with namd. *J Comput Chem*, **26**(16), 1781–1802.

Roh, C. (2012). A facile inhibitor screening of sars coronavirus n protein using nanoparticle-based rna oligonucleotide. *Int J Nanomedicine*, **7**, 2173.

Wishart, D. S. *et al.* (2006). Drugbank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res*, **34**(suppl_1), D668–D672.

Ziebuhr, J. *et al.* (2000). Virus-encoded proteinases and proteolytic processing in the nidovirales. *J Gen Virol*, **81**(4), 853–879.