Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Exploring new antiviral targets for influenza and COVID-19: Mapping promising hot spots in viral RNA polymerases

Inês Figueiredo-Nunes a, João Trigueiro-Louro a, b, ***, Helena Rebelo-de-Andrade a, b, *

a Host-Pathogen Interaction Unit, Research Institute for Medicines (iMed.ULisboa), Faculty of Pharmacy, Universidade de Lisboa, Av. Professor Gama Pinto, 1649-003, Lisbon, Portugal
b Antiviral Resistance Lab, Research & Development Unit, Infectious Diseases Department, Instituto Nacional de Saúde Doutor Ricardo Jorge, IP, Av. Padre Cruz, 1649-016, Lisbon, Portugal

ARTICLE INFO

Keywords:
- Influenza A virus
- Betacoronaviruses
- SARS-CoV-2
- PB1 protein
nsp12 protein
Conservation score
Druggability score
Consensus druggable pocket

ABSTRACT

Influenza and COVID-19 are infectious respiratory diseases that represent a major concern to public health with social and economic impact worldwide, for which the available therapeutic options are not satisfactory. The RdRp has a central role in viral replication and thus represents a major target for the development of antiviral approaches. In this study, we focused on Influenza A virus PB1 polymerase protein and the betacoronaviruses nsp12 polymerase protein, considering their functional and structural similarities. We have performed conservation and druggability analysis to map conserved druggable regions, that may have functional or structural importance in these proteins. We disclosed the most promising and new targeting regions for the discovery of new potential polymerase inhibitors. Conserved druggable regions of putative interaction with favipiravir and molnupiravir were also mapped. We have also compared and integrated the current findings with previous research.

1. Introduction

Influenza A virus (IAV) and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) are the etiologic agents of influenza and Coronavirus Disease 2019 (COVID-19), respectively. IAV and SARS-CoV-2 represent a major concern to public health and economy worldwide, because of their epidemic and pandemic potential, associated to unsatisfactory pharmacological prophylaxis and treatment strategies (Prioritizing, 2022). Current vaccines for influenza and COVID-19 are safe, although several limitations are widely recognized (National Institutes of Health, 2019). The influenza vaccine needs to be annually reviewed and their effectiveness might be compromised due to an antigenic mismatch between the strains included in the vaccine and the circulating strains (Lewnard and Cobey, 2018). Regarding COVID-19, the vaccination program for SARS-CoV-2 remains challenging from a global perspective, and there is still lack of robust information concerning possible seasonality and the duration and effectiveness of vaccinal-induced immunity against newly emerging SARS-CoV-2 variants (Baker et al., 2021). Regarding clinically authorized antiviral strategies, only a few options are available for treatment (National Institutes of Health, 2019). Moreover, usually the effectiveness of influenza antivirals decreases after the initial 48h of the symptoms’ onset (Naesens et al., 2016). Also, available drugs can become less effective due to the emergence of viruses with reduced susceptibility to antivirals (López-Medrano et al., 2012). Thus, the search for robust targets and new antiviral strategies directed against such targets is of utmost relevance. Antivirals currently represent the most reliable therapeutic strategy in the advent of a new pandemic virus, considering the presumable incapability to develop new vaccines for global use to control the early fast-spreading phase of the pandemic disease (Longini et al., 2005). Moreover, the existence of several antiviral options for influenza and COVID-19, with different mechanisms of action, might allow the combination of therapeutic strategies which can be more effective and minimize the selection of resistant viruses (Perelson et al., 2012).

IAV and SARS-CoV-2 are ribonucleic acid (RNA) viruses that contain an RNA-dependent RNA polymerase (RdRp) complex to express their proteins and replicate their genome (Hillen et al., 2020; Fodor, 2013).
extension; BR: Beta ribbon region; BH: Beta-hairpin region; CTE: C-terminal extension. Fingers, palm and thumb subdomains contain the polymerase motifs (not essential to form a suitable substrate for nuclear import, and consequently a functional trimeric RdRp (Fodor and Smith, 2004). While all viral titer that plausibly minimize disease transmission and severity (Hutchinson et al., 2011). Nuclear import of PA-PB1 dimers occurs due to the interaction between a nuclear import factor, the Ran binding protein 5 (RanBP5) and the nuclear localization sequences (NLS) of PB1 (Deng et al., 2006). The PA-PB1 dimer and the PB2 import from the cytoplasm, separately (Peng et al., 2020). The two copies of nsp8 and the nsp7 cofactors (CoV-RdRp) (Kirchdoerfer and Ward, 2019). It contains a right-hand-like structure with fingers, palm, and thumb subdomains, similarly to the structure described for the PB1 protein (Gao et al., 2020). Structure and domain organization of nsp12 are depicted in more detail in Fig. 1.

The interactions between the PB1, PA and PB2 subunits are crucial to maintain the IAV-RdRp structure and assure the normal functioning of the transcription and replication processes (Walker and Fodor, 2019). For that reason, the PA, PB1 and PB2 subunits of the IAV-RdRp complex, as well as the regions of interface between these subunits may represent interesting targets for drug discovery or design. The active site of IAV-RdRp is composed by the motifs A to F placed in the fingers, palm and thumb subdomains of the PB1 subunit that, together with the N-terminal region of the PB2 subunit, form a large central cavity (Pflug et al., 2017). Favipiravir is a RdRp inhibitor that targets the IAV-RdRp and blocks IAV genome copies replication (Pandey et al., 2020). However, its use is only approved in Japan and France for limited clinical conditions (Puruta et al., 2013a). Additionally, resistance molecular markers - namely, mutations in a highly conserved region of the IAV-RdRp (motif F) - have been already selected in vitro, which pose as a potential threat for antiviral resistance development (Goldhill et al., 2018).

The RdRp complex of the betacoronaviruses (beta-CoVs) is composed of the catalytic subunit non-structural protein 12 (nsp12) and the nsp8 and nsp7 cofactors (CoV-RdRp) (Kirchdoerfer and Ward, 2019). It contains a right-hand-like structure with fingers, palm, and thumb subdomains, similarly to the structure described for the PB1 protein (Gao et al., 2020). Structure and domain organization of nsp12 are depicted in more detail in Fig. 2.

Nsp12 subunit contains the CoV-RdRp active site, that is essential for both transcription and replication processes (Hillen et al., 2020). Similarly to the IAV-RdRp, the protein active site is formed by the conserved CoV-RdRp motifs A to G, in the fingers and palm subdomains of the nsp12 protein. These motifs contain essential residues for RNA synthesis, that interact and stabilize the RNA template. Specifically, motif B contains a loop region which, together with motif G residues, undergo conformational changes in response to bind the template-product (Yin et al., 2020). nsp8 and nsp7 are accessory subunits, that bind to nsp12 through the fingers (nsp8) and thumbs (nsp8 and nsp7) subdomains, respectively (Peng et al., 2020). The two copies of nsp8 and the nsp7 cofactors stabilize the nsp12 structure during the replication process and

**Fig. 1.** Schematic representation of the structure of IAV-RdRp PB1 protein. The PB1 protein has a characteristic right-hand fold (common to several RNA polymerases) and is composed by fingers, palm and thumb subdomains. Additionally, the protein is composed of a N-terminal extension (residues 1–34) and a C-terminal extension (residues 669–757), which interact with the PA and PB2 subunits, respectively. The interdomain limits are labeled with residues numbers. NTE: N-terminal extension; BR: Beta ribbon region; BH: Beta-hairpin region; CTE: C-terminal extension. Fingers, palm and thumb subdomains contain the polymerase motifs (not described in the picture). Coloured web version. If accepted, a black-and-white version of the figure will be provided for print.
Fig. 2. Schematic representation of the structure of CoV-RdRp nsp12 protein. The nsp12 (residues S1-Q932) is the core component of the complex and is composed of an N-terminal nidovirus RdRp associated nucleotidyltransferase (NiRAN) domain (residues D60-R249), an interface domain (residues A250-R365), and a nsp12-RdRp domain (residues S367–F920). The interdomain limits are labeled with residues numbers. Highly conserved polymerase motifs A-F are coloured in grey. NiRAN: N-terminal nidovirus RdRp associated nucleotidyltransferase; RdRp: RNA-dependent RNA polymerase. The RdRp domain contains a right-hand fold with fingers, palm and thumb subdomains

Coloured web version. If accepted, a black-and-white version of the figure will be provided for print.

I. Figueiredo-Nunes et al.
Virology 578 (2023) 45–60

prevent the premature dissociation of the tetrameric CoV-RdRp complex (Yin et al., 2020). A new direct-acting antiviral, molnupiravir, targeting the CoV-RdRp complex was recently approved by the U.S Food & Drug Administration (FDA) and the Medicines and Healthcare products Regulatory Agency (MHRA) for the treatment of COVID-19 (U.S Food and Drug Administration, 2022). Molnupiravir is a broad-spectrum antiviral that induces RNA mutagenesis during viral replication (Gordon et al., 2021). However, recent studies showed that molnupiravir has a less-than-expected effectiveness in decreasing hospitalization due to COVID-19 (Kozlov, 2021). Furthermore, the use of molnupiravir is only approved in the United States (US) and United Kingdom (UK) and exclusively for cases of mild-to-moderate disease in adults with risk factors for developing severe illness (U.S Food and Drug Administration, 2022; Medicines and Healthcare products Regulatory Agency, 2022).

Despite the opposite polarity of the viral genome and the low homology between IAV-RdRp and CoV-RdRp complexes (Buonaguro et al., 2020), there share some common general features. For instance, both contain a right-hand-like structure, with fingers, palm and thumb subdomains, that integrate the highly conserved motifs of the protein active site responsible for RNA interaction (Yin et al., 2020; Stubbs and Te Velthuis, 2014). Additionally, despite the conformational changes that occur in each IAV-RdRp and CoV-RdRp complexes, the PB1 protein and the nsp12 protein remain practically static, while the major rearrangements take place in the respective accessory subunits (te Velthuis and Fodor, 2016; Wang et al., 2020). Considering the analog function of these proteins and their structural similarities and, moreover, the existence of drug molecules as favipiravir and molnupiravir - that are known to interact with both RdRp complexes of these two viruses (Magro, 2020) – , it is important to map the pockets in these polymerases structures as the starting point to identify promising targets for each virus.

The development of a promising antiviral strategy primarily requires a drug target which is demanded to be (i) essential to the viral replication and (ii) druggable, i.e., it can interact and be modulated by potential inhibitors (Hajduk et al., 2005). Druggability is the protein’s ability to bind drug-like molecules with high affinity and specificity (Hopkins and Groom, 2002). Druggable pockets are characterized by physicochemical parameters like high volume, large depth and enclosure and a higher percentage of charged residues and hydrophobicity. Also, the size of the pocket should be >10 residues to enhance the interaction between the target and the ligand. “Decoy pockets” (pockets with less than 14 residues), may be incapable to properly bind a ligand (Hussein et al., 2015; Volkamer et al., 2012). Conservation studies allow to identify promising functionally/structurally important residues of the protein, that may represent therapeutically relevant disease-modifying targets (Blundell, 1996). Also, antiviral strategies targeting these residues are presumably less likely to become ineffective due to a mutation of the virus into a drug-resistant profile.

A robust structure-based analysis of the PB1 (influenza) and the nsp12 (beta-CoVs) polymerase proteins as drug targets has the potential to promote and accelerate the rational discovery of new antivirals, that target the most promising regions regarding the selected targets (Blundell, 1996; Kirchmair et al., 2011). Here, we focused on the PB1 protein from IAV-RdRp and the nsp12 protein from CoV-RdRp as drug targets for a prospective antiviral strategy. We studied the conservation and druggability patterns of these proteins, individually, to identify and thoroughly explore conserved druggable hot spots. Similar structure-based antiviral research studies were successfully performed in the recent past concerning the non-structural protein 1 (NS1) from IAV and the Spike protein from a wide range of beta-CoVs (Trigueiro-Louro et al., 2019, 2020, 2022; Darapameni et al., 2009; Sugiyama et al., 2009). Our findings highlight potential regions of functional or structural importance in these proteins, which may represent key targets for a potential pharmacological modulation of polymerase-mediated viral infection.

2. Results

For the conservation studies, each amino acid site within the PB1 subunit (IAV-RdRp) and the nsp12 protein (CoV-RdRp) sequences was scored between 0 (high variable) and 11 (highly conserved). As detailed in the Material and Methods, residues with conservation score of 11 corresponded to 100% amino acid identity; residues with conservation score (cs) ≥ 10 were considered highly conserved; residues with cs of 7–9 were considered conserved; and residues with cs ≤ 6 were considered variable. Residues with a cs of 11 were named top-conserved residues. Druggable residues/regions were predicted and scored from 0 to 2 (highly druggable). Residues with the maximum druggability score of 2 in all crystallographic structures were named top-druggable residues. Residues/regions both top-conserved and top-druggable were named top-ranked hot spots and represent the most promising sites to be considered for the development of a prospective antiviral strategy targeting either the IAV-RdRp or the CoV-RdRp.

2.1. Influenza PB1 protein

2.1.1. Conservation studies

Considering the influenza PB1 protein: 64.60% of the residues are top-conserved; 94.32% are conserved (cs between 7 and 9); and only 5.68% are considered variable (cs ≤ 6). Cs of each residue from the different PB1 domains are described in Table 1. Variable residues are mostly located at the fingers (namely, residues: T57, V113, G154, G216) and thumb subdomains (namely, residues: Q584, A587, R621, S741). Apart from variable residues, fingers subdomain contains 87.89% highly conserved residues (cs between 10 and 11); while thumb subdomain contains 88.81% (which represents 119 residues from a total of 134 that comprise this domain, as shown in Table 1). The palm subdomain showed the most promising results regarding the conservation analysis, with 94.12% highly conserved residues, 85.00% of which are top-conserved. Altogether, this resulted in three regions with the maximum cs: E78-A93 (from fingers subdomain), G274-M290 and V439-A453 (from palm subdomain). The N-terminal and C-terminal extensions – that interact with PA and PB2 subunits, correspondingly, – are vastly conserved, showing highly conserved/variable residues rates of 94.29%/5.71% and 90.91%/9.09%.
### Table 1
Conservation grade classification for each amino acid position on the IAV PB1 protein.

Residues with conservation score of 11 have the highest grade corresponding to absolute conserved sites (100% amino acid identity; top-conserved residues); residues with conservation score ≥10 were classified as highly conserved; residues with conservation score of 7–9 were classified as conserved; and residues with conservation score ≤6 were classified as variable.

NTE: N-terminal extension; BR: Beta ribbon region; BH: Beta-hairpin region; CTE: C-terminal extension.

Coloured web version. If accepted, a black-and-white version of the figure will be provided for print.

| Region | Conservation score ≤6 | Conservation score 7–9 | Conservation score ≥10 |
|--------|-----------------------|------------------------|------------------------|
| NTE    | A14                   | V12                    |                        |
| BR     | E177, M179, M195, Q210| I181, V200, R211       |                        |
| Fingers| Q48, R52, T57, E75, V113, G154, T156, D175, L212, N213, G216, R237, I339, A401 | K54, L108, M111, V114, V149, M171, E172, M174, K176, R215, T257, S261, M317, I336 | G35, T36, G37, Y38, T39, M40, D41, T42, V43, N44, R52, T57, E75, V113, G154, T156, D175, L212, N213, G216, R237, I339, A401 |
|        |                       |                        |                        |
| Palm   | L298, E457, V473, K486| K430, 433, H456, D464  |                        |
|        |                       |                        |                        |
| BH     | R361, L364, S375, E383, K387, I397 | I368, M372, S384, R386 |                        |

(continued on next page)
Table 1 (continued)

| Region | Conservation score ≤6 | Conservation score 7-9 | Conservation score ≥10 |
|--------|------------------------|------------------------|------------------------|
| Thumb  | IS17, LS76, QS84, AS87, PS96, DS87, D618, D619, R521, E538, S67 |
|        | IS25, S573, D581, K586, L587, E598, S678 |
|        | R489, Y499, G500, F501, V502, A503, N504, F505, S506, M507, E508, L509, P510, S511, F512, G513, V514, S515, G516, S518, E519, S520, A521, S522, M523, S524, G56, V527, T528, V529, T530, K531, N532, N533, M534, E535, N536, N537, D538, L539, G540, P541, A542, T543, A544, Q64, Q545, G546, A547, L548, Q549, L550, F551, S552, K553, D554, Y555, R556, Y557, T558, Y559, R560, G561, S562, R563, G564, D565, T566, G567, E568, Q569, S570, R571, R572, F573, S574, K575, S576, L577, E578, E579, Q582, I583, N584, S585, G586, S587, L589, V590, S591, S592, D593, G594, S595, N596, N597, G598, Y599, N600, I601, R602, N603, L604, H605, I606, P607, E608, V609, M610, L611, K612, W613, E614, L615, K616, D617, Y618, G619, S620, E621, R622, L623, L624, G625, N626, P627, N628, N629, P630, F631, V632, S633, H634, K635, E636, I637, S638, V639, Y640, A641, Q642, Y643, V644, P645, E646, H647, G648, L649, G650, P651, A652, K653, M654, E655, N656, L657, R658, S659, Q660, E661, T662, T663, H664, N665, W666, P667, R668, Q669, S670, R671, C672, L673, E674, G675, D676, S677, L678, L679, R680, C681, E682, L683, D684, D685, E686, M687, G688, V689, Q690, C691, G692, N693, H694, Q695, D696, E697, D698, P699, T700, R701, S702, S703, S704, T705, S706, R707, P708, G710, T711, S712, S713, M714, V715, E716, A717, M718, V719, S720, R721, A722, R723, I724, D725, A726, R727, D729, F730, T731, T732, G733, R734, I735, K736, K737, E738, F740, E742, I743, M744, K745, S747, T749, I750, E751, L753, R754, R755, Q756, K757

Table 1 (continued)

| Region | Conservation score ≤6 | Conservation score 7-9 | Conservation score ≥10 |
|--------|------------------------|------------------------|------------------------|
| PL     | V645, S654             | N642, M646             |
|        | V640, N641, A643, V644, P645, E648, H649, G650, P651, A652, K653, M654, E655, N656 |

respectively. The priming loop region of the PB1 structure, known as essential for the replication cycle of the IAV-RdRp, is composed of 18 amino acids, from which 14 are highly conserved and the remaining 4 have lower cs: N642 (cs = 9.48), V645 (cs = 5.50), M646 (cs = 8.38) and S654 (cs = 4.78).

2.1.2. Druggability studies

For the druggability analysis, we selected structures of the IAV-RdRp PA-PB1-PB2 trimers: 6QNW, 6QPF and 6TUS. We identified 209 druggable pockets with the DoGSiteScorer (DGSS) webserver along with 449 top-druggable residues in within the highly promising druggable pockets. These were mostly located in fingers subdomain (127 residues), followed by thumb subdomain (120 residues) and palm subdomain (92 residues). Hence, several regions with the maximum druggability score can be highlighted, namely: N283-D295, T489-A521 and M523-P541. The priming loop region was not crystallized in none of the analyzed structures. When considering the SARS-CoV-2 sequences of IAV-RdRp (I-CDP), the C-terminal regions comprised 27 and 45 highly druggable residues, respectively, that correspond to 79.41% and 50.56% of the total residues from these regions. We additionally explored the most promising conserved druggable regions (three-dimensionally close), which formed CDPs, as described in the Materials and Methods section. We identified 10 CDP with ≥14 aa in the PB1 structure of IAV-RdRp (I-CDP): The described I-CDP are mostly placed in the fingers (I-CDP6, I-CDP8, I-CDP10), palm (I-CDP3, I-CDP4, I-CDP5, I-CDP6), and thumb subdomains (I-CDP1, I-CDP2, I-CDP3, I-CDP4, I-CDP5) of the influenza PB1. All I-CDP are described in Table 2 and the top five pockets with higher sizes were mapped onto the PB1 tridimensional structure (PDB entry: 6QNW; Fig. 3).

Some residues were spotted in more than one I-CDP, namely: F22, I484, F496, Q569, T570, R572. These represent both highly druggable and highly conserved residues (cs ≥10). The integration of conservation and druggability patterns for the several domains of the PB1 protein structure of IAV-RdRp is described in Table S1 (Supplemental data).

2.2. Beta-CoV nsp12 protein

2.2.1. Conservation studies

2.2.1.1. SARS-CoV-2. Overall, we found that the nsp12 protein of SARS-CoV-2 is highly conserved. When considering the SARS-CoV-2 sequences – that include SARS-CoV-2 ancestral lineage and all the SARS-CoV-2 emerging variants – 97.42% of the residues are highly conserved (cs ≥10), 2.25% are conserved (cs ≤7–9) and 0.32% are variable (cs ≤6). The N-terminal extension of nsp12 (formed by NiRAN, β-hairpin and interface domains) contains only 0.55% variable residues, while the remain 94.79% (which represents 346 residues) are highly druggable residues. The nsp12-RdRp domain (composed by fingers, palm, and thumb subdomains) exhibits the most promising results with respect to conservation analysis, with 85.82%; 85.14%; and 82.05% top-druggable residues, respectively. From the nsp12-RdRp domain, there were only spotted 4 residues with cs ≤9: Y521 (cs = 9.48), V645 (cs = 5.50), M646 (cs = 8.38) and S654 (cs = 4.78).
nsp12 protein sequences from SARS-CoVs (that includes both SARS-CoV-2 and SARS-CoV) we noticed a slightly lower level of conservation, mostly in residues from the NiRAN and interface domains. Considering the complete structure of the nsp12 protein, the percentage of highly conserved residues decreases from 94.79% to 96.35% and the number of variable residues (cs ≤ 6) increases from 3 to a total of 20 (2.14%). Residue G671 shows a cs of 0 in all groups referred so far. Palm subdomain is the most conserved domain, with 99.42% top-conserved residues. Fingers subdomain remains highly conserved (comparing with previous analysis of SARS-CoV-2), with 99.27% highly conserved residues. Despite the changes in conservation rates, there are also many large sequence regions of top-conserved residues, namely V398–S520, E522–A580, L630–G670 (from fingers subdomain), T687–E811 (from palm subdomain) and P830–A923 (from thumb subdomain).

2.2.1.3. SARSr-CoVs and MERS-CoV. Sequences from the beta-CoVs included in the study (namely, SARSr-CoVs and MERS-CoV) were also comparatively analyzed. In this group, the conservation pattern changed considerably. Sequence conservation of each position of SARS-CoV-2, SARS-CoV, and SARSr-/MERS-CoVs nsp12-RdRp domain is presented in detail in Table 3.

Once again, the nsp12-RdRp domain stands as the most conserved region, particularly the thumb subdomain. SARS-CoV-2, SARS-CoV and SARSr-/MERS-CoVs sequences all show 86.05% top-conserved residues for this subdomain. Despite the overall slightly lower percentage of highly conserved residues, we have identified 619 top-conserved residues. This resulted in multiple regions comprised by top-conserved residues: S635–C646 (fingers subdomain) Q698–S709, K751–C765, K783–M794 (palm subdomain), D618–N628, S672–G683, (fingers and palm subdomains), A840–T853, A878–H898, E917–Q932 (thumb subdomain) and E802–K821 (palm and thumb subdomains).

2.2.1.4. Druggability studies. The available structures of SARS-CoV-2 and SARS-CoV RdRp complexes were collected and four of them were selected (PDB IDs: 6M71, 6NUR, 6NUS and 7BTF), according to the strategy described in Materials and Methods.

Table 2 Consensus Druggable Pockets (CDPs) identified for the PB1 protein tridimensional structures of influenza A virus. Pockets were identified according to their size, on a descendent order from 1 (larger) to 10 (smaller). Only pockets with ≥10 residues were considered. Positions placed in the PB1 subunit of each consensus pocket are distributed according with the respective PB1 domains. I-CDP: influenza consensus druggable pockets. Coloured web version. If accepted, a black-and-white version of the figure will be provided for print.
nsp12 subunit monomers or in association with nsp8 and nsp7 accessory subunits (tetramers). NiRAN domain showed the lowest druggability score in both nsp12 monomers (19.38% druggable residues) and tetramers (25.55% druggable residues). 

β-hairpin and interface regions each showed approximately 50% residues with low druggability scores. Residues of β-hairpin region were only resolved for SARS-CoV-2 nsp12 structures. Besides its low druggability rate, the interface region is comprised by 2 top-druggable regions: V318–F324 and S346–H350. 

Globally, within these three domains, we only identified 40 top-druggable residues. The nsp12-RdRp domain presents promising druggability results. In this domain, we have identified 155 new druggable residues pertaining to druggable pockets with the maximum druggability in SARS-CoV-2 and SARS-CoV monomers. Similarly, 224 highly druggable aa were pointed-out for tetramers. In each conformation, the identified residues are mostly part of fingers subdomain. Furthermore, 140 top-druggable residues of these are shared by both conformations and are mainly from the palm subdomain. Some of these residues are spatially close and form CDPs.

3. Discussion

Structure-based approaches targeting highly conserved and highly druggable regions for the discovery and development of more robust antivirals are lacking (Trigueiro-Louro et al., 2020, 2022). The RdRp has a central role in viral replication (Choi, 2012). Likewise, and considering its importance, it represents an antiviral target by excellence in many infectious diseases. The available anti-polymerase drugs approved for treatment of IAV and SARS-CoV-2 infections, such as favipiravir and molnupiravir, have shown limited effectiveness, so a thoroughly study of the IAV-RdRp and the CoV-RdRp potential for the development of alternative anti-polymerase strategies – or ways to optimize the existent ones – is of utmost importance. In the current work, we mainly explored the IAV-RdRp (with a major focus on the PB1 protein) and the CoV-RdRp (mainly regarding the nsp12).

3.1. Influenza PB1 protein

Overall, we identified 297 top-ranked hot spot residues (considering the entire structure of the protein), which were mostly placed in fingers (82 residues) and palm (76 residues) subdomains. It can be readily hypothesized that targeting top-ranked hot spot residues within these subdomains could impact the viral RNA synthesis.

Likewise, the N-terminus and C-terminus of the PB1 showed promising results in our analysis, with the identification of 19 and 24 top-
Table 3
Overall alignment of SARS-CoV-2, SARSr-CoV and MERS-CoV sequence conservation patterns of nsp12-RdRp domain. Analysis of nsp12 sequences from SARS-CoV-2, SARSr-CoV (SARS-CoV-2 and SARS-CoV) and SARSr-/MERS-CoV group at each amino acid position from the RdRp-domain according to its conservation. The druggable sites were predicted according to the results of two independent bioinformatic tools: DGSS and SF. Each site is coloured according to the conservation score scale (0-11) showed in this figure. Highly variable residues are coloured in red (scores 0–3), variable residues in pink (scores 4–6), conserved residues in light blue (scores 7–9) and highly conserved residues in dark blue (scores 10–11). Highly conserved, conserved, and variable residues represented 79.94%, 9.12% and 10.94%, respectively, considering all sequences in study. Distinct conservation patterns are mostly present in NiRAN domain, in which it was reported a decrease of the cs in 49.8% residues. Nevertheless, 68 were top-conserved in all grouped sequences, namely: S1, F7, L8, R19, V202, G203, I205-Y217, F219, G220, D221, F222, V234, Y237, Y238, S239, L240, L241, M242, I244, L245, T246.
Our studies showed that all these residues are highly conserved (except levels of viral messenger RNA (mRNA) detected (Sugiyama et al., 2009). Single V715S mutation also resulted in a replication with a reduced level of viral RNA (vRNA) replication in vitro. Moreover, the process (Sugiyama et al., 2009). Specifically, Sugiyama et al. showed that V715/V750 and I746/V750 double mutants showed a significantly lower level of viral RNA (cRNA) for initiation of vRNA synthesis (Fan et al., 2019). Our studies showed that this domain bears one of the highest percentages of variable residues; notwithstanding, it also presented promising conserved druggable residues in specific regions, as previously highlighted. Targeting conserved druggable residues on this subdomain, such as the ones identified in Table S1, could result in an antiviral strategy that endangered the viral replication cycle by inhibiting protein dimerization. Particularly, residues K553, Y557 and R571, that were described to interact with the 5-terminal nucleotides of cRNA (Fan et al., 2019), have consistently promoted (i) conformational changes that allow the opening of IAV-RdRp binding site and (ii) the correct positioning of complementary strands. Especially, residues K553, Y557 and R571 were crucial for protein dimerization; and consequently promotes (i) conformational changes that allow the opening of IAV-RdRp binding site and (ii) the correct positioning of complementary strands. Particularly, residues K553, Y557 and R571, that were described to interact with the 5-terminal nucleotides of cRNA (Fan et al., 2019), have cs ≥ 10 and are also top-druggable residues. Also, we

## Table 4

| Consensus pocket | Position (and corresponding nsp12 regions) | nsp12-RdRp domain | Size (aa) |
|------------------|--------------------------------------------|-------------------|----------|
|                  | NIRAN | Interface | Fingers | Palm | Thumb |
| C-CDP1           | 478 479 480 482 483 579 | 582 583 584 585 586 587 588 | 30 |
|                  | 131 132 165 172 | 175 176 243 244 245 | 246 247 248 249 | 250 251 252 316 319 | 320 | 460 461 462 463 630 | 701 787 788 789 791 | 29 |
| C-CDP2           | 388 390 392 400 402 404 405 406 407 | 408 446 447 448 449 542 545 546 547 | 28 |
|                  | 298 299 300 301 302 | 303 304 307 | 568 571 572 63 62 63 65 636 639 640 | 648 649 651 652 654 655 658 659 662 | 663 | 24 |
| C-CDP3           | 333 342 343 344 345 | 356 358 359 360 363 | 20 |
|                  | 164 165 166 172 | 131 132 165 172 | 589 591 592 597 600 601 604 | 605 607 696 746 749 750 753 | 754 755 756 | 19 |
| C-CDP4           | 452 453 456 540 542 553 554 556 557 | 558 560 621 623 624 665 667 676 | 19 |
|                  | 458 459 460 462 621 | 625 626 627 | 620 791 792 793 794 795 798 | 685 689 | 19 |
| C-CDP5           | 429 430 433 434 435 | 436 350 351 352 353 354 355 356 357 | 15 |
|                  | 231 232 233 234 235 | 289 290 291 292 305 | 467 470 471 474 | 730 731 732 733 734 735 736 | 737 738 742 | 15 |
| C-CDP6           | 369 370 373 374 370 | 530 534 535 536 537 | 15 |
|                  | 365 358 359 360 363 | 141 142 | 122 126 129 130 133 | 134 135 138 139 140 | 141 142 | 416 441 548 | 840 841 843 844 | 15 |
| C-CDP7           | 452 453 456 540 542 553 554 556 557 | 558 560 621 623 624 665 667 676 | 15 |
|                  | 164 165 166 172 | 164 165 166 172 | 589 591 592 597 600 601 604 | 605 607 696 746 749 750 753 | 754 755 756 | 15 |
| C-CDP8           | 458 459 460 462 621 | 625 626 627 | 620 791 792 793 794 795 798 | 685 689 | 13 |
| C-CDP9           | 495 497 498 500 501 | 502 512 516 560 | 561 562 565 569 573 576 577 580 | 17 |
|                  | 490 491 493 494 495 | 749 584 591 592 593 594 595 596 597 598 599 | 12 |
| C-CDP10          | 547 548 549 551 552 553 554 555 | 12 |

Coloured web version. If accepted, a black-and-white version of the figure will be provided for print.

Regarding the PB1 thumb domain, this region was previously described by Fan et al. as crucial for protein dimerization; and consequently promotes (i) conformational changes that allow the opening of IAV-RdRp binding site and (ii) the correct positioning of complementary RNA (cRNA) for initiation of vRNA synthesis (Fan et al., 2019). Our studies showed that this domain bears one of the highest percentages of variable residues; notwithstanding, it also presented promising conserved druggable residues in specific regions, as previously highlighted. Targeting conserved druggable residues on this subdomain, such as the ones identified in Table S1, could result in an antiviral strategy that endangered the viral replication cycle by inhibiting protein dimerization. Particularly, residues K553, Y557 and R571, that were described to interact with the 5-terminal nucleotides of cRNA (Fan et al., 2019), have cs ≥ 10 and are also top-druggable residues. Also, we
identified residues K553 and Y571 as part of the I-CDP1 (Table 5). Fan et al. also described that the 3′ cRNA interacts with the IAV-RdRp through the PB1 loop (residues F551-R571) and the PA residues E300, Y464, K488 and R496 (Fan et al., 2019). It is worth to mention that we also identify the PA residues Y464, K488 and R496 as part of the I-CDP1. The development of a potential inhibitor targeting this pocket might impact on the stability of the interaction between the RdRp and the cRNA. Also, the same binding site of 3′ cRNA was identified in the RdRp of influenza B virus (Fan et al., 2019). Future studies that include structures of polymerase from other influenza virus types should be considered to identify common consensus pockets that could accommodate, for instance, a pan-polymerase inhibitor that could be active against both influenza A and influenza B viruses.

As previously explained, it is important to identify target regions that have an important role in protein function to be able to discover a potential inhibitor that could, effectively and significantly impact the viral replication process (Mandal et al., 2009). For instance, the interface of interaction between RanBP5 and the PB1 subunit is essential for nuclear import of the PA-PB1 complex from the cytoplasm and consequently further assembly with PB2 to form a functional trimeric IAV-RdRp complex (Deng et al., 2005). Moreover, previous studies reported that the knock-down of the RanBP5-PB1 interaction resulted in lower levels of vRNA and nuclear PA-PB1 dimers detection, in vitro (Deng et al., 2006). RanBP5-binding site is exclusively placed in the first N-terminal 299 aa of the PB1 subunit, and it includes residues from its bipartite NLS such as K188, R189, K208 and K209 (Deng et al., 2006). These residues were already described as conserved and with a potential role in providing a viral fitness advantage (Hutchinson et al., 2011). Our studies are consistent with the high cs previously identified for this region, although a low druggability score has been attributed (Table 5). Notwithstanding, the overall PB1 N-terminal region (M1-S299) shows promising top-ranked hot spots and I-CDP3, I-CDP4, I-CDP6, I-CDP8 and I-CDP9 are placed in this region (Table 2). An antiviral strategy directed against this protein region could not only target an early step of the viral replication, by interfering in the nuclear import of the PA-PB1 complex but might also disturb the interaction between PA and PB1 subunits that occurs in both the cytoplasm and the nuclear compartments of the infected cells.

The binding site of favipiravir ribofuranosyl-5′-triphosphate (F-RTP), the active form of favipiravir, with the H3N2 polymerase protein was also described by Sada et al. It includes the following PB1 residues: K121, Q124, M246, Y253 and T257 (Sada et al., 2020). Our studies show that these residues are highly druggable (except T257 which showed a cs of 7). These residues are also characterized by high druggability scores (except M246; druggability score = 0). Residues K121, Q124 and Y253 are also part of the I-CDP9 identified in the study, which is placed in the highly promising fingers subdomain of the PB1 protein (Table 5).

Goldhill et al. and Abdelnabi et al. highlighted the importance of PB1 motif F of the PB1 protein for the antiviral mechanism of action of favipiravir (Goldhill et al., 2018; Abdelnabi et al., 2017). Motif F is placed in fingers subdomain of the PB1 subunit and comprises residues T223-P244. According to our analysis, these are top-conserved residues, (excluding R233; cs = 9.96); and the residues T223, L224, N225, T226, T243 and P244 are also highly druggable (Table 5). Considering that the PB1 crystallographic structures available in the protein databases are not fully resolved for this region, it was not possible to properly study its druggability potential. Future analysis should include the motif F of PB1, considering its prospective importance as an antiviral target (Sheahan et al., 2020). Furthermore, considering the reported interaction of favipiravir with both IAV-RdRp and CoV-RdRp, a more comprehensive study of the motif F may potentially lead to the identification of additional pockets with common features in different RNA polymerases (e.g., PB1 and nsp12), and consequently provide insights for the development of an antiviral strategy directed against both SARS-CoV-2 and influenza virus.

3.2. Beta-CoV nsp12 protein

Our conservation studies showed that the nsp12 is highly conserved throughout its domains regarding different beta-CoVs (Table 3), even when considering the newly emerged Omicron variant. This suggests that this protein may display lower evolutionary rates compared with more-exposed surface proteins. In this vein, one can hypothesize that an antiviral or immunotherapeutic approach focusing on the CoV-RdRp may represent a promising strategy for the prevention and/or treatment of COVID-19, with lower chances of becoming ineffective due to the inherent viral diversity and evolution. Regarding the comparative study of the conservation and druggability analysis, we have identified 47 top-ranked hot spot residues in the N-terminal extension of nsp12 and 164 top-ranked hot spot residues in the nsp12-RdRp domain, mainly in the fingers subdomain. These represented top-druggable residues in both protein conformations and top-conserved residues in the distinct groups of viral sequences analyzed (Table 3) and so exhibit a higher antiviral target potential. Ultimately, C-CDP such as C-CDP1, C-CDP7, C-CDP8, C-CDP9 and C-CDP13 should be the focus for a prospective antiviral strategy targeting the nsp12-RdRp domain. These may represent the starting point for the structure-based discovery of new molecules with potential to inhibit the nsp12 functions or destabilize the protein conformational structure.

Our studies highlighted the fingers and palm subdomains as the most promising regions for a future development of antiviral molecules that target SARS-CoV-2, SARSr-CoVs and MERS-CoV nsp12. These subdomains contain the highly conserved motifs that are responsible to mediate template-directed RNA synthesis (Hillen et al., 2020). Interestingly, molnupiravir was described to bind to fingers subdomain, including residues K545, A547, I548, S549, R555 (Sheahan et al., 2020). In our studies, we have identified these residues as top-ranked hot spots and as part of the 17-residues-long C-CDP10, which is fully placed in the highly promising fingers subdomain and contains other catalytic residues from the nsp12 protein motif F. A summary table integrating results collected from our studies and available data collected from experimental studies performed by other authors and published in the literature regarding nsp12 residues is described in Table 6, similarly to the one performed for the IAV-RdRp PB1 protein.

Gao et al. previously explored the SARS-CoV-2 RdRp and identified residues in highly conserved motifs which were also highly conserved in RNA viruses as hepatitis C virus and poliovirus (Gao et al., 2020). The most promising residues reported by Gao et al. included: K545, R553 and R555, which are part of the nucleoside triphosphate (NTP) entry channel (motif F); the divalent-cation-binding residue D618 and the classic SDD region of motif C formed by residues S759, D760 and D761. Residues K545 and R555 are included in the molnupiravir binding site and the results from our study are consistent with both (i) their structural/functional relevance for the nsp12 protein and (ii) their suitable characteristics for a possible interaction with an antiviral molecule. The remaining residues described by Gao et al. (R553, D618, S759, D760 and D761) are top-conserved residues (in every group of sequences analyzed) but only the R553 and D761 are also top-druggable residues. It is also worth to highlight that the C-CDP10 identified in our study includes the residue R553 (Table 6) (relevant residue previously reported by Gao et al. (2020). Considering the mechanism of action of the molnupiravir, it is acknowledgeable that targeting some key residues belonging to these regions will impair SARS-CoV-2 replication (Kabinger et al., 2020). Furthermore, considering the potential of a polymerase inhibitor (or the optimization of molnupiravir) focused on a larger number of these highly promising top-ranked hot spots may prompt a more effective inactivation of the CoV-RdRp active site and consequently the blockade of the RNA-template binding.

Wang et al. reported that the molecular rearrangements that occur in the nsp12 protein during the replication cycle are mainly determined by two of the highly conserved motifs: B (residues G682 to T686) and G (residues D499 to L514) (Schueler-Furman and Baker, 2003). The results
from our study are consistent with the structural relevance of these residues, showing that they are mainly highly conserved across SARS- and MERS-CoVs nsp12 sequences; motif B is also highly druggable in both nsp12-conformations. On the other hand, motif G has only three residues with high druggability scores (K500, S501, G502), from which K500, and S501 were described to interact with the RNA-template (Schueler-Furman and Baker, 2003). The residues K500, S501 and G502 are placed in CDP9 identified by our study, which also includes part of motif B (key residue: T685) (Table 6). The top-ranked hot spots found within these regions might allow the discovery of an anti-nsp12 inhibitor, which can interfere with the conformational determinants of the polymerase protein and consequently, affect the viral replication.

There are antiviral drugs which have been reported to inhibit the function of both IAV-RdRp and CoV-RdRp, such as favipiravir (Qi et al., 2021; Furuta et al., 2013b). This suggests that RdRp complexes have key common structural determinants that allow the interaction with the same ligands for the inhibition of their activity. Hence, it is reasonable to hypothesize that there can be other common/consensus pockets that allow the binding of a putative molecule that can pharmacologically modulate the polymerase function or interfere with its structure. For instance, favipiravir has been described to bind to residues from the nsp12 protein of the CoV-RdRp active site, as K545, K551, R553, R555, D618, K621, D761 and K798 (Naydenova et al., 2020). These residues are placed in the fingers and palm subdomains from the nsp12 protein and, which were previously described to have suitable conservation and druggability patterns – highly promising target regions (Table 6). Most of these residues comprise the C-CDP10 (K545, K551, R553 and R555), which is also composed by residues that were described to interact with molnupiravir. Considering the ability of favipiravir to bind to IAV and SARS-CoV-2, this consensus pockets could represent the starting point for the design of an antiviral molecule with the ability to inhibit the function of both IAV-RdRp and CoV-RdRp and for the development of a broad-spectrum (or even universal) antiviral strategy.

This study represents a computational approach to study both the PB1 protein from IAV-RdRp and the nsp12 protein from CoV-RdRp as promising antiviral targets. However, experimental approaches need to be performed to provide us additional insights on regions with crucial roles on IAV and SARS-CoV-2 polymerases structure and/or function. We suggest that further studies should be mainly focused on the top-ranked hot spots highlighted in the fingers and palm subdomains of PB1 and nsp12 proteins. Also, a structure-based rational strategy for drug discovery/design of new inhibitor compounds – or the optimization of the authorized/in approval drugs – requires the study of the effects of pharmacological blockade through the highly promising CDPs that we identified in our study. We suggest that further studies should be mainly focused on the top-ranked hot spots highlighted in the fingers and palm subdomains of PB1 and nsp12 proteins.

4. Conclusion

It is important to discover new and more effective antiviral strategies that could contribute to overcome the current widely recognized therapeutic limitations and to be readily available for prevention and treatment of influenza and COVID-19. This is even more relevant considering the diversity and unpredictable evolution patterns of IAV and beta-CoVs.

In this work, we performed conservation and druggability studies focusing on the IAV PB1 protein and the beta-CoVs nsp12 protein, to identify promising hot spots which can later provide the basis for the structure-based discovery/design of anti-polymerase compounds. The combination of the conservation and druggability analysis disclosed 10 CDPs and 158 top-ranked hot spots for the PB1 protein of IAV-RdRp and 17 CDPs and 164 top-ranked hot spots for the nsp12 protein of CoV-RdRp – which were mostly placed in the fingers and palm subdomains of these proteins. Also, in this study we highlighted I-CDP9 and C-CDP10, which are not only composed by residues already described to bind favipiravir and/or molnupiravir, respectively, but also contain other residues described for its importance in polymerase structure and function.

Fig. 4. Mapping results of the five highest-size and grade consensus druggable pockets onto the nsp12-polymerase protein three-dimensional structure. Ribbon diagram of nsp12-polymerase monomeric chain (PDB entry 6M71) are showed in (A) two side views related by 180-degree rotation; (B) top view and (C) bottom view related by 90-degree rotation. The consensus druggable pockets are coloured as follows: C-CDP1, green; C-CDP2, blue; C-CDP3, orange; C-CDP4, purple; and C-CDP5, yellow. Coloured web version. If accepted, a black-and-white version of the figure will be provided for print.
function. Furthermore, I-CDP1, I-CDP3, I-CDP4, I-CDP6, I-CDP8 and I-CDP10 (from PB1 structure) and C-CDP1, C-CDP7, C-CDP9 and C-CDP13 (from nsp12 structure) are also highly promising pockets composed by conserved druggable residues which potentially have a relevant role on proteins’ functions and structure during viral replication. Focusing on these CDPs can provide the structural basis for the discovery/design of new and potentially more effective polymerase inhibitors – via its pharmacological modulation -, or the optimization of the authorized/in more ambitious perspective, this study may also provide preliminary efficiency and consequent overcome current antiviral limitations. In a

Table 5
Summary table of promising IAV-RdRp PB1 residues explored in this study and respective experimental findings described in literature regarding their potential role on protein function/structure. Conservation score scale: 0–11. Druggability score scale: 0–2.

| Residue | Domain | Conservation/Druggability scores | I-CDP | Findings | Authors |
|---------|--------|---------------------------------|-------|----------|---------|
| K715    | C-     | 9.73/2                          | –     | Double   | Sugiyama et al. (Sugiyama et al., 2009) |
| 746     | Terminal | 5.25/0                          |       | mutations |                     |
| V750    | Extension | 11/1                           |       | 715/750 and 746/750 were associated with lower levels of vRNA in vitro. |                     |
| K553    | Thumb   | 11/2                            | I-CDP1 | Residues reported to be involved in IAV-RdRp | Fan et al. (Fan et al., 2019) |
| Y557    |        | 11/2                            | I-CDP5 | Dimerization and interaction with 3’ cRNA. Binding site of 3’ cRNA identical in Influenza B virus |                     |
| R571    |        | 9.98/2                          | I-CDP1 | –         |                     |
| K188    | Beta-ribbon | 11/0                          | –     | Residues from the bipartite NLS that interact with RanBP5 (mediate the nuclear import of the PA-PB1 complex). | Hutchinson et al.; Deng et al., 2006; |                     |
| R189    |        | 11/9                            |       | –         |                     |
| K208    |        | 9.78/0                          |       | –         | Sugar et al. (Sugiyama et al., 2009) |
| K209    |        | 9.97/1                          |       | –         |                     |
| K121    | Fingers | 9.98/2                          | I-CDP9 | Binding-site of the active form of favipiravir. | Sada et al. (Sada et al., 2020) |
| Q124    |        | 11/2                            | I-CDP9 | –         |                     |
| M246    |        | 11/0                            | –     | –         |                     |
| Y253    |        | 9.89/2                          | I-CDP9 | –         |                     |
| T257    |        | 7.98/2                          | –     | Residues from motif F of the PB1 protein. | Goldhüll et al.; Abdelnabi et al., 2017 |
| T223    | Fingers | 11/2                            |       | –         |                     |
| L234    |        | 11/2                            |       | –         |                     |
| N225    |        | 11/2                            |       | –         |                     |
| T226    |        | 11/2                            |       | –         |                     |
| T243    |        | 11/2                            |       | –         |                     |
| P244    |        | 11/2                            |       | –         |                     |

5. Materials and Methods

5.1. Conservation studies

5.1.1. Dataset construction and sequence analysis

Regarding the Influenza polymerase protein conservation studies, 984 PB1 nucleotide sequences of worldwide circulating viruses of human-infecting Influenza A subtypes – H1N1, H2N2, H3N2, H5N1 and H7N9 – were obtained from GISAID’s EpiFlu™ database at www.gisaid.org (Shu and McCauley, 2017; Elbe and Buckland-Merrett, 2017). For the conservation studies of the nsp12 protein of worldwide circulating human beta-CoVs, sequences from SARS-CoV-2 variants – Alpha, Beta, Delta, Gamma, Lambda, Mu and Omicron available until 6th December (regarding omicron) were collected from GISAID’s EpiCov™ database (www.gisaid.org). Sequences of the nsp12 protein of SARS-CoV-2 ancestral lineage (until March 24, 2020), SARS-CoV and MERS-CoV were also collected from public databases, including: GISAID’s EpiCov™ database (www.gisaid.org) (Shu and McCauley, 2017; Elbe and Buckland-Merrett, 2017), GenBank (https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/) (Sayers et al., 2018), 2019 Novel Coronavirus Resource (2019nCoVR) (https://bigd.big.ac.cn/ncov) (Zhao et al., 2020), NIAID Virus Pathogen Resource (ViPR) database (https://www.viprbrc.org) (Pickett et al., 2012) and NCBI Virus database (https://www.ncbi.nlm.nih.gov/labs/virus) (Hatcher et al., 2017). A total of 3019 nsp12 nucleotide sequences (nucleotide positions 13442–16236 of the Orf1ab gene) were collected. All available sequences with a complete coding region were selected for the study. Duplicated sequences and sequences containing degenerate nucleotides were removed. The final datasets, which were further analyzed, contained 926 PB1 nucleotide sequences (influenza A virus) and 2890 nsp12 nucleotide sequences (beta-CoVs). The sequence alignment was performed using the Multiple Alignment using Fast Fourier Transform (MAFFT) available at https://www.ebi.ac.uk/Tools/multix/mafft/ (Katoh, 2002; Madeira et al., 2019). Aligned nucleotide sequences were then automatically translated into their respective amino acid sequences using MEGA X (https://www.megasoftware.net/). Using the translation command, the program automatically extracts all protein-coding domains for translation into their respective amino acid sequences using the genetic code – predefined settings and displays the corresponding protein sequence (Kumar et al., 2018).

5.1.2. Conservation analysis

Position-specific conservation scores of the MSA were calculated using the Valdar scoring method from the Jalview AACons Web server (version 2.11.1.4) (Waterhouse et al., 2009; Valdar, 2002). This scoring method strives to normalize the data against redundancy without losing any evolutionary data (Valdar, 2002). Conservation studies of NS1 protein of influenza A virus and of Spike protein sequence of SARS-CoV-2, have been previously achieved using this algorithm (Triqueiro-Louro et al., 2019, 2022; Choi, 2012; Darapeniet al., 2009). Conservation degree was calculated for each residue position in the MSA and a conservation threshold of 7 was established (Livingstone and Buckland-Merrett, 1993). Residues with conservation score of 11 corresponded to 100% amino acid identity; residues with conservation score ≥10 were considered highly conserved; residues with conservation score of 7–9 were considered conserved; and residues with conservation score ≤6 were considered variable.

5.2. Druggability studies

5.2.1. Crystallographic structures selection

Crystallographic structures of the PB1 subunit of Influenza A polymerase and the nsp12 of SARS-CoV-2 polymerase protein were obtained from the RCSB Protein Data Bank (PDB), a freely repository available through the website www.rcsb.org (Berman et al., 2000). Reasonable crystallographic systems for polymerase protein were explored and

5.3. IAV-RdRp function analysis

5.3.1. Characterization of functional residues

5.3.2. Analysis of conserved residues

5.4. Summary

6. Concluding remarks

6.1. Future directions

6.2. Acknowledgments

6.3. References

7. Appendices
potential druggable pockets on the protein surface, based only on the druggability score scale: 0–11. Druggability score scale: 0–2.

5.2.2. Druggability prediction

The druggable sites of the chosen crystallographic structures were predicted, – for either monomeric, dimeric and tetrameric forms – from a consensus of two distinct studies: the webserver DoGSiteScorer (DGSS) and the commercial software MOE-SiteFinder (SF) (Volkamer et al., 2012; Chemical Computing Group ULC. MOE (Molecular Operating Environment), 2013). The third approach initially used, the webserver PockDrug (PD), was further withdrawn, considering that the selected structures exceeded the size limit set by this specific server. Both programs required the upload of a structure in PDB format.

5.2.3. Druggability study design

The druggability prediction for each polymerase-conformation of influenza and SARS-CoV-2 viruses was independently studied, in a first stage. Tridimensional structures were divided in two conformational groups: (1.1) dimers (PDB IDs: Z2NL, Z2TT, 3A1G and 3CM8) and (1.2) tetramers (PDB IDs: 6QNF, 6QPF and 6TUS) were used for the druggability studies. Regarding beta-Cov, crystallographic structures of the nsp12-12p-12p-7 polymerase complex from SARS-CoV-2 (6M71 and 6NUR) and SARS-CoV (6NUS and 7BTF) were selected for study. For all crystallographic 3D structures: i) all crystallized water molecules, bound ligands and other coordinated molecules were removed from the PDB files; ii) the protonation states of the amino acid residues were assigned, and energy minimization was performed using integrated tools from the MOE 2015.10001 program (Chemical Computing Group ULC. MOE (Molecular Operating Environment), 2013).

5.2.2. Druggability prediction

The druggable sites of the chosen crystallographic structures were predicted, – for either monomeric, dimeric and tetrameric forms – from a consensus of two distinct studies: the webserver DoGSiteScorer (DGSS) and the commercial software MOE-SiteFinder (SF) (Volkamer et al., 2012; Chemical Computing Group ULC. MOE (Molecular Operating Environment), 2013) DoGSiteScorer server was able to predict potential druggable pockets on the protein surface, based only on the geometric (size, shape) and physicochemical descriptors (grid-based method). Pockets labeled as druggable had a druggability score ≥0.4 (on a scale between 0 and 1) (Peng et al., 2020). SiteFinder of MOE 2015.10001 software uses a geometry-based method based on parameters as the receptor size and the number of contacts with the receptor; and considers the chemical properties information to predict potential pockets for drug target. The top ranked sites with a positive score, were considered as “druggable” (Chemical Computing Group ULC. MOE (Molecular Operating Environment), 2013). The third approach initially used, the webserver PockDrug (PD), was further withdrawn, considering that the selected structures exceeded the size limit set by this specific server. Both programs required the upload of a structure in PDB format.

### Table 6

| Residue | Domain | Conservation/Druggability scores | C-CDP | Findings | Authors |
|---------|--------|----------------------------------|-------|----------|---------|
| K545    | Fingers| 11/2                             | C-CDP10 | Interaction with molnupiravir. | Seahan et al. (Seahan et al., 2020) |
| A547    | 11/2   |                                  |       |          |         |
| I548    | 11/2   |                                  |       |          |         |
| S549    | 11/2   |                                  |       |          |         |
| R555    | 11/2   |                                  |       |          |         |
| K545    | Fingers| 11/2                             |       | Residues from the NTP entry channel (motif F) | Gao et al. (Gao et al., 2020) |
| R553    | 11/2   |                                  | C-CDP10 |          |         |
| D618    | Palm   | 11/2                             | C-CDP17 | Divalent cation-binding residue | Gao et al. (Gao et al., 2020) |
| S759    | Palm   | 11/0                             |       | SDD region (motif G) | Gao et al. (Gao et al., 2020) |
| D760    | 11/1   |                                  |       |          |         |
| D761    | 11/2   |                                  | C-CDP17 |          |         |
| K500    | Fingers| 11/0                             | C-CDP9 | Essential for nsp12 protein conformational rearrangements (motif G) | Schauer-Furmer et al. (Schauer-Furmer and Baker, 2003) |
| S501    | 11/2   |                                  |       |          |         |
| G502    | 11/2   |                                  |       |          |         |
| K545    | Fingers| 11/2                             | C-CDP10 | Favipiravir binding site in the RdRp active site | Naydenova et al. (Naydenova et al., 2020) |
| K551    | 10/2   |                                  |       |          |         |
| R553    | 11/2   |                                  |       |          |         |
| D618    | Palm   | 11/2                             | C-CDP17 |          |         |
| K798    | 11/2   |                                  | C-CDP17 |          |         |
score ≥1 (sites identified in more than half of the total of structures considered by both algorithms) were considered druggable for each conformation. The druggability consensus were integrated and analyzed in parallel with the conservation information to identify druggable pocket/sites and residues that overlap or are spatially close to the conserved polymerase regions/residues. Moreover, for each polymerase complex, we unified the pocket druggability information from DGSS server to identify regions located spatially closed, which could form larger pockets. These were named “consensus druggable pockets” (CDPs). Only CDPs present in over half of the 3D crystallographic structures in analysis, for each virus protein-conformation, and formed by sites with the same expression were considered, to guarantee meaningful results. Likewise, only pockets with ≥10 amino acid (aa) residues were selected.

**Funding**

This work was supported by the grant PD/BD/128402/2017 from Fundação para a Ciência e a Tecnologia (FCT) PhD Programme in Medicines and Pharmaceutical Innovation (i3DÚ).

**Credit author statement**

Inês Figueiredo-Nunes: Methodology, Formal analysis, Investigation, Writing (Original Draft), Writing (Review & Editing), Visualization, Joao M. Trigueiro-Louro: Conceptualization, Methodology, Investigation, Formal analysis, Writing (Original Draft), Writing (Review & Editing), Helena Rebelo-de-Andrade: Conceptualization, Resources, Writing (Review & Editing), Supervision, Project administration, Funding acquisition.

The authors would like to thank to all authors, originating and submitting laboratories that share their sequences through the GISAID screening techniques. Infect. Disord.: Drug Targets 11, 64–69. https://doi.org/10.1016/j.virol.2022.11.001.

**Declaration of competing interest**

The authors declare no conflict of interest.

**Acknowledgments**

The authors would like to thank to all authors, originating and submitting laboratories that share their sequences through the GISAID EpiFlu®, Influenza Research Database, NCBI Influenza Virus, GISAID’s EpiCov®, GenBank, 2019 Novel Coronavirus Resource (2019nCoVR), NIAID Virus Pathogen Resource (ViPR) and NCBI Virus databases.

The authors also thank Rita Guedes Lab Group (Medicinal Chemistry Unit, iMed.ULisboa, FFUL) not only for providing the required computational equipment but also for sharing their expertise regarding some computer methodologies applied to this study.

**Appendix A. Supplementary data**

Supplementary data to this article can be found online at https://doi.org/10.1016/j.virol.2022.11.001.

**References**

Abdelnabi, R., Mozais, A.T.S., Lysen, P., Imbert, I., Beauchart, S., Blanc, H., Froyen, M., Vignuzzi, M., Canard, B., Neys, J., Delang, L., 2017. Understanding the mechanism of the broad-spectrum antiviral activity of favipiravir (T-705): key role of the FI motif of the viral polymerase. May 26 J. Virol. 91 (12). https://doi.org/10.1128/JVI.00487-17, e00487-17. PMID: 28391577; PMCID: PMC5446660.

Baker, R.E., Park, S.W., Wagner, C.E., et al., 2021. The limits of SARS-CoV-2 mutagenesis via the RNA template. Jul J. Biol. Chem. 297 (1), 100770. https://doi.org/10.1074/jbc.R120.016676.

Berman, H.M., Batistuzzi, B., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., et al., 2000. The protein Data Bank. Nucleic Acids Res (Internet) 28 (1), 235–242. https://doi.org/10.1093/nar/28.1.235. Available from: Blundell, T.L., 1996. Structure-based drug design. Nov J. Nature 384 (6604 Suppl. 1), 25–26. https://doi.org/10.1038/384023a0. PMID: 8895997.
