



Predicted structural mimicry of spike receptor-binding motifs from highly pathogenic human coronaviruses

Christopher A. Beaudoin^{a,*}, Arian R. Jamasb^{a,b}, Ali F. Alsulami^a, Liviu Copoiu^a, Andries J. van Tonder^c, Sharif Hala^{d,e}, Bridget P. Bannerman^a, Sherine E. Thomas^a, Sundeep Chaitanya Vedithi^a, Pedro H.M. Torres^f, Tom L. Blundell^{a,*}

^a Department of Biochemistry, Sanger Building, University of Cambridge, Tennis Court Rd, Cambridge CB2 1GA, United Kingdom

^b Department of Computer Science & Technology, University of Cambridge, JJ Thomson Ave, Cambridge CB3 0FD, United Kingdom

^c Department of Veterinary Medicine, University of Cambridge, Madingley Rd, Cambridge CB3 0ES, United Kingdom

^d King Abdullah International Medical Research Centre – Ministry of National Guard Health Affairs, Jeddah, Saudi Arabia

^e King Saud bin Abdulaziz University for Health Sciences, Jeddah, Saudi Arabia

^f Laboratório de Modelagem e Dinâmica Molecular, Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil

ARTICLE INFO

Article history:

Received 19 May 2021

Received in revised form 26 June 2021

Accepted 27 June 2021

Available online 02 July 2021

Keywords:

SARS-CoV-2

SARS-CoV

MERS-CoV

Coronavirus spike protein

COVID-19

Viral host mimicry

Infectious disease

Structural bioinformatics

ABSTRACT

Viruses often encode proteins that mimic host proteins in order to facilitate infection. Little work has been done to understand the potential mimicry of the SARS-CoV-2, SARS-CoV, and MERS-CoV spike proteins, particularly the receptor-binding motifs, which could be important in determining tropism and druggability of the virus. Peptide and epitope motifs have been detected on coronavirus spike proteins using sequence homology approaches; however, comparing the three-dimensional shape of the protein has been shown as more informative in predicting mimicry than sequence-based comparisons. Here, we use structural bioinformatics software to characterize potential mimicry of the three coronavirus spike protein receptor-binding motifs. We utilize sequence-independent alignment tools to compare structurally known protein models with the receptor-binding motifs and verify potential mimicked interactions with protein docking simulations. Both human and non-human proteins were returned for all three receptor-binding motifs. For example, all three were similar to several proteins containing EGF-like domains: some of which are endogenous to humans, such as thrombomodulin, and others exogenous, such as *Plasmodium falciparum* MSP-1. Similarity to human proteins may reveal which pathways the spike protein is co-opting, while analogous non-human proteins may indicate shared host interaction partners and overlapping antibody cross-reactivity. These findings can help guide experimental efforts to further understand potential interactions between human and coronavirus proteins.

© 2021 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Viruses have long been known to utilize molecular mimicry of host proteins to interrupt and exploit host biochemical pathways during infection [1,2]. Alongside the need to employ host machinery for the viral replication cycle, the evolution of viral protein motifs that resemble host proteins can result in new virulence mechanisms, such as inducing inflammation and evading the immune system [3]. Coronaviruses, in particular, have been suspected to have acquired human protein mimics throughout the long record of human coronavirus infections [4,5]. As further evi-

dence, the highly pathogenic human coronaviruses, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), SARS-CoV, and the Middle Eastern Respiratory Syndrome Coronavirus (MERS-CoV), have been shown to encode numerous short linear motifs across their genomes that are homologous to human proteins [6]. Although coronavirus infections are typically localized to the lungs, resulting in respiratory infections, viral material has also been found in other organs, such as the kidney, brain, and heart, resulting in more life-threatening infections [7]. Furthermore, SARS-CoV-2 (the causative agent of the COVID-19 pandemic) infection has presented symptoms not previously seen in other coronavirus infections, such as conjunctival discharge from the eyes [8,9]. Investigations into coronavirus host mimicry may shed light on viral tropism and infection severity [10].

* Corresponding authors.

E-mail addresses: cab233@cam.ac.uk (C.A. Beaudoin), tlb20@cam.ac.uk (T.L. Blundell).

The structure of the receptor-binding motif (RBM) on the spike glycoprotein is particularly important for determining the tropism of the virus [11]. Host receptors that contain motif(s) that complement the electrochemical and spatial configurations of the viral RBM will interact and, thus, initiate viral entry [12,13]. Angiotensin converting enzyme II (ACE2) has been established as the primary cell entry receptor for SARS-CoV-2 and SARS-CoV and dipeptidyl peptidase IV (DPP4) as the primary cell entry receptor for MERS-CoV. However, several reports, some preliminary, have proposed additional coronavirus cell entry receptors, such as transferrin receptor protein 1, kidney injury molecule-1, kremen protein 1, and α v integrins for SARS-CoV-2 [14–20]. Additionally, coronavirus spike proteins have been proposed to interact with host factors to facilitate infection aside from their role in cell entry [21]. For instance, two studies found that the SARS-CoV-2 spike protein alone can interact with the blood brain barrier [22,23]. The importance in receptor-binding and low glycosylation surrounding the coronavirus RBM residues make it an attractive target for inhibition by small-molecule drugs, therapeutic peptides, and neutralizing antibodies [24–26].

To date, there has been limited investigation into the structural similarity of highly pathogenic coronavirus RBMs [27]. Identifying structurally analogous human proteins may give insight into endogenous biochemical pathways that the virus is hijacking to facilitate infection or may help explain autoimmune disorders triggered by coronavirus infections [28,29]. Detecting similar microbial proteins may reveal shared host receptors or antibody cross-immunity [30]. Short linear motifs on coronavirus spike RBMs have been shown to share high amino acid sequence identity with human proteins, which may indicate host mimicry [31–34]. However, protein structure and fold similarity have been shown as more informative than amino acid sequence similarity in predicting molecular mimicry [35,36]. Drayman *et al.* performed a structural similarity search using bacterial and viral motifs and experimentally validated the simian vacuolating virus 40 major capsid protein mimicry of Gas6 binding with TAM – Tyro3, Axl, and Mer – receptors, demonstrating that structural paralogs with low amino acid identity may still act as molecular mimics. Thus, to add to the understanding of host mimicry of highly pathogenic coronavirus RBMs, we used structural bioinformatics tools to model and map the extent to which the three-dimensional structures of the SARS-CoV-2, SARS-CoV, and MERS-CoV spike RBMs are potentially mimicking the interactions of experimentally-determined protein structures. We used structural alignment tools with distinct methodologies to perform a structural similarity screen between the RBMs and all known protein structures and, subsequently, tested potential RBM interactions with protein–protein docking simulations. Several cell signaling proteins, innate immune factors, snake and spider toxins, and microbial antigens are found to share structural features with the three RBMs. This information may help guide experimental efforts to elucidate spike RBM interactions, including that of vaccine design and cell entry receptor discovery.

2. Results and discussion

2.1. Receptor-binding motif structural similarities and characteristics

Several models of the spike protein for each of the highly pathogenic coronaviruses have been experimentally determined; however, many of them are missing residues due to the difficulty in resolving the structure of flexible protein motifs [37]. To overcome this issue and obtain a representative three-dimensional model of each spike receptor-binding motif (RBM), we used ProtCHOIR, a

recently developed pipeline to automate the modelling of homo-oligomers, to model each trimeric spike protein and, subsequently, manually selected the RBM residues for each coronavirus (Fig. 1). All generated models were structurally aligned to experimental models using TM-align to determine modelling precision. On a scale from 0 to 1, a TM-score of over 0.5 between two proteins implies that they have the same fold, while below 0.2 suggests a random alignment. Each RBM alignment with the corresponding experimental structure reported a TM-score over 0.95, reflecting high-quality modelling. Although receptor-binding of coronavirus spike proteins has been shown to be an elaborate process that involves interactions with glycans and multiple protein domains, we selected the most interactive region of the spike RBD with primary receptors (i.e. ACE2 for SARS-CoV and SARS-CoV-2; DPP4 for MERS-CoV) from experimental models as the receptor-binding motif (RBM) [38].

The structural similarity of the RBMs to one another was quantitatively assessed using TM-align before assessing their similarity to other known proteins. The SARS-CoV-2 and SARS-CoV RBMs were very similar with a TM-score of 0.71, while the TM-scores of MERS-CoV with the other two were both less than 0.25 (Fig. 1). This level of divergence is also reflected at the amino acid sequence level for the RBD of SARS-CoV-2 and SARS-CoV at 64.6% sequence identity and MERS-CoV with SARS-CoV-2 and SARS-CoV at 19% and 21.6%, respectively.

As seen in Fig. 1, the SARS-CoV-2 RBM is comprised almost exclusively of hydrophobic and polar non-charged amino acids, with the exception of one acidic glutamate and one basic lysine. SARS-CoV is similar to SARS-CoV-2 in that it is composed mostly of hydrophobic and polar non-charged residues with some exceptions as single amino acid differences, such as an acidic aspartate in the middle of the SARS-CoV RBM. The MERS-CoV RBM consists of more acidic and basic amino acids and contains fewer polar non-charged residues. Of note, the SARS-CoV-2 and SARS-CoV RBMs have 7 and 8 aromatic residues, respectively, exposed on the receptor-binding surface of the RBM. The recent discovery of the N501Y and E484K mutants add a potentially functional aromatic and basic residue, respectively, in the SARS-CoV-2 RBM – both of which have been proposed to increase binding to ACE2 [39]. Modelling of the mutants yielded very small structural changes in the SARS-CoV-2 RBM – TM-scores of the mutant RBMs aligned to the reference structure were above 0.9.

In terms of global architecture, the SARS-CoV-2 and SARS-CoV RBMs contain two anti-parallel beta-strands connecting three loops, although the SARS-CoV-2 RBM has two short beta-strands leading to a cystine disulfide loop (Fig. 1). Both SARS-CoV and SARS-CoV-2 contain a similar cystine disulfide bond helping shape one end of the respective RBMs. The MERS-CoV RBM consists of three beta-strands connecting four loops. Because loop flexibility may affect overall structure, we submitted each RBD to the CABS-flex 2.0 web server and found that the cystine disulfide loop of both the SARS-CoV-2 and SARS-CoV RBMs displayed high flexibility (>9 RMSF) – otherwise, the RBM residues on all three RBMs were predicted to exhibit low RMSF (less than 6.5) (Supplementary Fig. 1). The flexibility predictions from CABS-Flex 2.0 were supported by separate studies on coronavirus RBMs [40,41]. The high flexibility of the cystine loops in the SARS-related RBMs motivated the use of two additional models provided by CABS-Flex 2.0 for the structural similarity screen. The added models reported surprisingly low TM-scores compared to the references (0.42 and 0.65 for SARS-CoV-2 and 0.45 and 0.41 for SARS-CoV), revealing the high flexibility in these loops (Supplementary Fig. 1). Overall, SARS-CoV-2 and SARS-CoV were found to share higher structural homology with one another than in comparison with MERS-CoV.

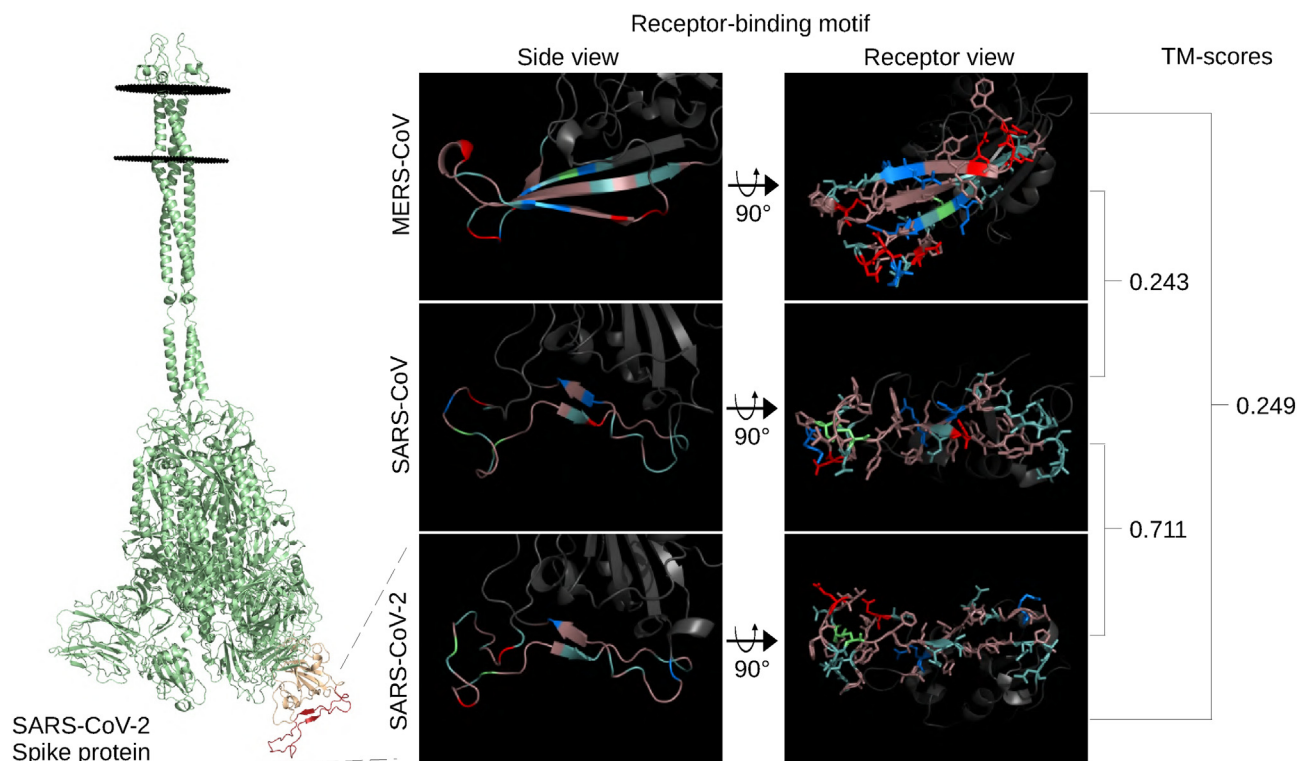


Fig. 1. Spike receptor-binding motif comparison. The full-length SARS-CoV-2 spike protein (green), left, modelled using ProtCHOIR is shown with the receptor-binding domain (yellow) and receptor-binding motif (red) marked. The RBMs from the side view are shown, middle, with the amino acids labelled by color: red for acidic (D,E), blue for basic (H,R,K), light teal for polar non-charged (S,N,T,Q), dirty violet for hydrophobic (A,V,I,L,M,F,W,P,G,Y), and lime green for cysteine residues. RBMs from the host cell receptor side are shown, right, with amino acid stick configurations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.2. Structural similarity screen

After RBM model generation, we performed a structural similarity screen for each RBM. Four sequence-independent 3D-structure alignment tools with different methodologies were used to quantify the structural similarity between the RBMs and known 3D protein structures in order to better understand shared structural features between the RBMs and potential mimics. Notably in this study, although spike may engage in interactions within human cells, we focused on protein structures that would be found in the extracellular matrix (excluding antibodies, due to their structural diversity) to gain more insight into potential cell entry receptors, immunopathies, and shared antigenicity with other microorganisms [42].

The PDBFold, RUPEE, and HMI-PRED web servers were used, and TM-align was locally-installed and run pairwise against the downloaded PDB database clustered at 100% sequence identity. The TM-score distributions between SARS-CoV-2 and SARS-CoV were quite similar, while MERS-CoV was more similar to a greater number of proteins (Fig. 2A). The MERS-CoV RBM returned 3,954 structures with a TM-score of over 0.5 (~top 1% of TM-scores) out of 245,055 total RBM-chain alignments and an average TM-score of 0.33. The SARS-CoV-2 and SARS-CoV RBMs had lower average TM-scores, 0.298 and 0.297 respectively, and the top 1% corresponded roughly to the 0.4 TM-score line. Thus, structures with a TM-score of > 0.4 were selected for further analysis for the SARS-related viruses: 4,025 for SARS-CoV-2 and 3,561 for SARS-CoV. PDBFold returned 621–806 and 1,163 structures for the SARS-related and MERS-CoV RBM models, respectively. The top 1,000 hits from each RUPEE run were recorded. HMI-PRED outputs ranged from 20 to 50 mimicked PDB templates per RBM. All alignments of interest were manually inspected to validate the

potential for structural mimicry. Returned aligned proteins from each tool were linked to their corresponding PDB and UniProt codes. Shared UniProt codes between two or more tools were regarded as high-confidence hits. Biologically relevant structural alignments specific to each tool were also inspected and considered. Structural alignments that would not make sense biologically, such as when the RBM is facing the inside of the protein, were discarded, while alignments that were logical but found outside of protein–protein interfaces were included on a case-by-case basis. Returned structures not shown to be found in the extracellular matrix were removed. All tools returned their respective spike structures, confirming their validity.

A total of 62 UniProt codes, excluding 28 toxins, were considered as biologically relevant, which were comprised of 35, 19, 19, and 8 selected alignments from RUPEE, HMI-PRED, PDBFold, and TM-align, respectively. When comparing tools (Fig. 2B), RUPEE and PDBFold web servers shared 7 UniProt codes for at least one RBM, while TM-align shared 1 with PDBFold and 0 with RUPEE. HMI-PRED shared 2 structures with RUPEE, 1 with PDBFold, and 0 with TM-align. Little overlap was shown between most of the tools, which is consistent with structural similarity-based studies on HIV and human proteins [43]. The combined returned UniProt codes, excluding toxins, from all four tools totalled 39, 23, and 29 for the SARS-CoV-2, SARS-CoV, and MERS-CoV RBMs, respectively. The top alignments consisted of cytokines, chemokines, and growth factors and their receptors; structures containing EGF-like domains; complement activation proteins; cystine disulfide-rich toxins derived from snakes and spiders; and antigenic microbial proteins. A Venn diagram showing some shared hits between the three RBMs can be seen in Figure 2C, and a full listing of the hits, alignment values, and tools can be found in Tables 1 and 2. The SARS-CoV and SARS-CoV-2 RBMs shared more

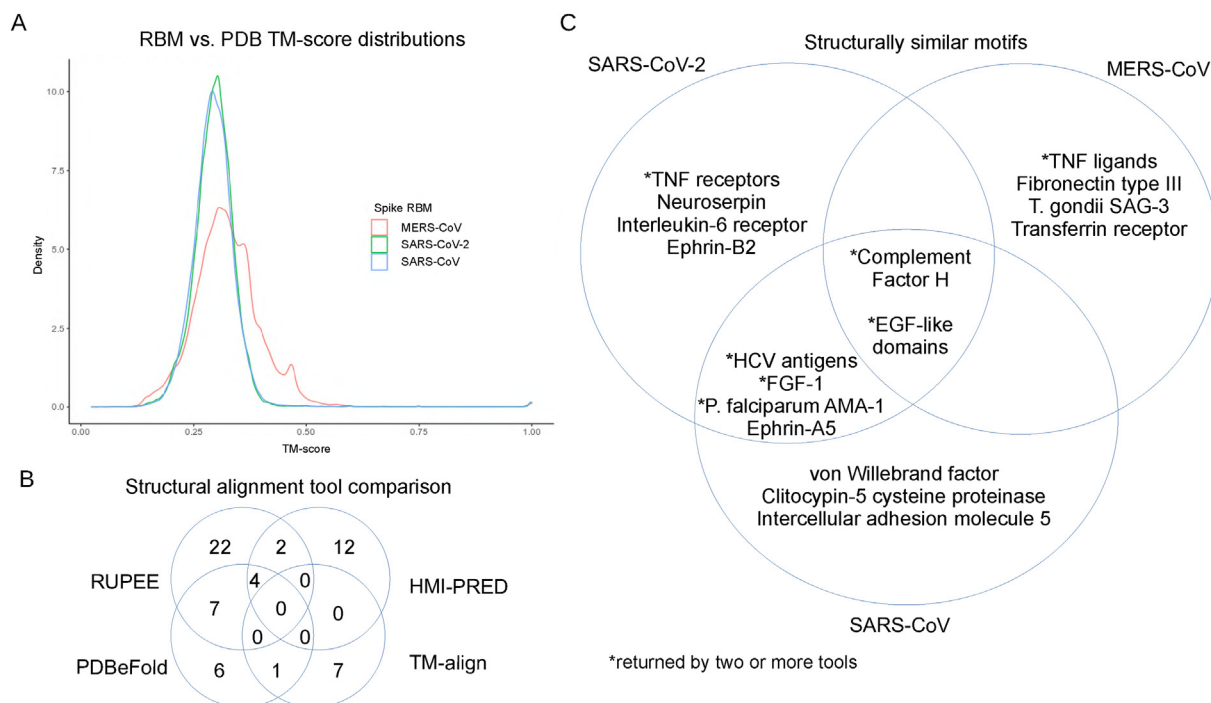


Fig. 2. Summary of structural similarity screen. The TM-scores generated from the in-house TM-align screen are displayed as a density plot for each RBM (A). Number of shared proteins from the tools used in the structural similarity screen compared and contrasted (B). Structurally similar motifs, common between coronavirus receptor-binding motifs, compared and contrasted (C).

structural domains, while MERS-CoV returned more unique hits compared to the other two. Altogether these results indicate that proteins from completely different protein families may interact with coronavirus spike RBMs.

2.3. Analysis of predicted structural mimicry

Further examination of the structural alignments and their relevance to biological activity was performed to elucidate potential mechanisms of molecular mimicry by the SARS-CoV-2, SARS-CoV, and MERS-CoV spike RBMs. The UniProt and STRING databases were used to link the predicted mimics with potential interaction partners, and the PDB provided template structures to determine whether the alignments were found in ligand-binding regions. Selected high-confidence potential interactions were further evaluated using protein–protein docking with ClusPro PIPER in order to better understand electrochemical, in addition to structural, complementarity considering the low amino acid sequence identity. The docked models were then analyzed with the FoldX AnalyseComplex program to determine the complex interaction energy. Docking of the natural ligand to the receptor was performed to obtain a control interaction energy. The energy of the original PDB protein complex was also predicted as an experimental control. The exploration of these interactions with structural alignment visualization and protein–protein docking may help explain their potential roles in infection.

The potential mimics were split into two categories: endogenous vs. exogenous, or human vs. non-human, to more effectively describe the results in the context of infection. Mimicry of endogenous proteins may reveal which human pathways, specifically, the viral RBM is hijacking; structurally similar exogenous proteins may exhibit shared interference of human interaction pathways or antigenicity with the coronavirus RBMs. Endogenous hits, both discovered by single and multiple structural alignment tools, are summarized in Table 1 and exogenous hits in Table 2.

2.4. Endogenous

Several proteins containing EGF-like domains were found to be similar to all three RBMs. EGF-like domains are evolutionarily conserved domains that share homology to the epidermal growth factor and have been shown to function primarily in tissue organization and repair [44,45]. Both the cystine disulfide loop and the central beta-strand sub-motif structures in the SARS-CoV-2 and SARS-CoV RBMs and the MERS-CoV beta-strands were found to mimic EGF-like domains.

The EGF-like domain of the urokinase-type plasminogen activator (uPa) in complex with its receptor, urokinase plasminogen activator receptor (uPAR), (PDB: 2fd6) was found to be similar to both SARS-CoV-2 and SARS-CoV RBMs using RUPEE. Interestingly, the uPa/uPAR system has been implicated in SARS-CoV-2 pathogenesis with uPAR as an early predictor of severe respiratory failure [46,47]. Although the RBMs protrude into the receptor in the structural alignments, the alignments suggest that the RBMs might bind to uPAR (Supplementary Fig. 2A).

The neurogenic locus notch homolog protein 1 (NOTCH1) EGF-like domain was returned for the SARS-CoV-2 RBM central beta-strands and MERS-CoV RBM by RUPEE and PDBeFold. NOTCH1 is involved in developmental, innate immunity, and inflammation signaling pathways, and natural ligands of the NOTCH1 EGF-like domains include jagged-1, jagged-2, delta-like 1 (DLL1), DLL3, and DLL4 [48]. Alignment of the SARS-CoV-2 and MERS-CoV RBMs with the EGF-like domain of NOTCH1 bound to DLL4 (PDB: 4x11) shows potential for molecular mimicry, i.e. the coronavirus RBMs may bind to DLL4 (Supplementary Fig. 2B) [49]. The SARS-CoV-2 RBM was also found similar to NOTCH2 by RUPEE, but no PDB complex models were available for further inspection. No direct interactions with the NOTCH1 pathway have been revealed, but its inhibition has been proposed to help fight SARS-CoV-2 infection [50].

All three RBMs were found to potentially mimic the EGF-like domain of coagulation factor VIIa. Further inspection of the

Table 1
Structural alignment values and data for endogenous hits.

Representative PDB			Mimic UniProt Name	Alignment Length	Alignment Sequence Identity (%)	TM-score	RMSD	Tools
MERS-CoV	SARS-CoV	SARS-CoV-2						
1dx5	1zaq	1mox 1dqb	Transforming growth factor alpha Thrombomodulin	25 21;20;25	4 4.8;5;4	0.25 0.31074;0.20501; 0.2719	3.53 2.76;3.02;3.08	R R,P and H (SARS-CoV-2)
5ky0		1toz	Neurogenic locus notch homolog protein 1	18;23	5.6;4.3	0.29958;0.26322	2.04;2.67	R,P
		5mwb	Neurogenic locus notch homolog protein 2	19	0	0.23683	2.81	R
1xka	2vwl	1fjs	Coagulation factor X	23;21	17.4;0	0.29888;0.24978	2.67;2.88	R,P
3th4	5ky2	1bf9	Coagulation factor VII	25;22	8;4.5	0.27601;0.25062	3.00;2.97	R,P
5f85		1edm	Coagulation factor IX	17;20	5.9;5	0.33731;0.25615	1.27;2.91	R,P
	3bt1	2fd6	Urokinase-type plasminogen activator	23	8.7	0.25522	3.13	R
		5wb7	Epiregulin	19;25	5.3;4	0.30636;0.23605	2.39;3.52	R
		5wb8	Epigen	22	13.6	0.34781	2.67	R
	1emn	1emn	Fibrillin	26;25;27	3.8;8;11.1	0.24425;0.26079	2.67;3.40;3.59	R
4hsv	1rhp	1nap	Platelet factor 4 (Platelet basic protein)	26;23;25	3.8;4.3;8	0.33673;0.19566; 0.28516	2.67;3.77;2.97	R, H and P (SARS-CoV-2)
	2r3z	1o7z	Small inducible cytokine B10	27;28	0;7.1	0.28217;0.30758	3.38;2.9	R
	3oj2	3cu1	Fibroblast growth factor 1	37;21	5.4;14.3	0.29384;0.27446	3.54;2.74	H,P
	4c2b		von Willebrand factor	31	3.2	0.27715	3.53	H
	1b3e	1b3e	Serotransferrin	54;50	7.4;10	0.42052;0.41965	3.86;3.83	T
	4l0p	4m4r	Ephrin-A5	39;37	10.3;2.7	0.27482;0.31208	4.36;3.68	H
		2wo2	Ephrin-B2	34	8.8	0.3341	3.08	H
	1b53	1b50	C-C motif chemokine 3	26;22	0;0	0.24893;0.21513	3.13;2.97	R
		2h62	Bone morphogenetic protein 2	32	3.1	0.28716	3.69	H
	4xfu	4r6u	Interleukin-18	31;46	0;2.2	0.33910;0.34459	2.88;4.37	R
1qnk			C-X-C motif chemokine 2	27	7.4	0.2786	3.53	P
1je4			C-C motif chemokine 4	24	8.3	0.32333	3.12	P
3qb4			Growth/differentiation factor 5	39	17.9	0.35895	3.78	H
5mw5			Jagged-2	38	15.8	0.54617	2.56	T
4xbm			Delta-like protein 1	37	5.4	0.51041	2.32	T
5mvx			Delta-like protein 4	38	5.3	0.50604	2.47	T
		5fuc	Interleukin-6 receptor subunit alpha	36	0	0.34661	2.87	R
		1pvh	Interleukin-6 receptor beta chain	34	5.9	0.32838	3.51	R
4nqc			TCR beta chain	28	10.7	0.40641	2.42	H
5t5w			Interferon lambda receptor 1	27	7.4	0.37707	2.65	H
		2hey	Tumor necrosis factor receptor superfamily member 4	32	6.2	0.31459	3.33	R
		1ncf	Tumor necrosis factor receptor superfamily member 1A	24	4.2	0.29911	2.94	R
		2aw2	Tumor necrosis factor receptor superfamily member 14	25	12	0.23837	3.42	R
		1oqe	Tumor necrosis factor receptor superfamily member 13C	23	4.3	0.25858	3.29	R,P
3v56			Tumor necrosis factor ligand superfamily member 13B	40	12.5	0.39446	3.31	H,R
4en0			Tumor necrosis factor ligand superfamily member 14	37	16.2	0.44547	2.86	H,R
1hfh	5o32	1hfi	Complement factor H	33;35;28	9.1;2.9;3.6	0.3487;0.31744; 0.27510	3.39;3.82;3.57	R,P (no SARS- CoV),H (SARS- CoV)
3oed			Complement receptor type 2 (CR2)	31	6.5	0.3175	3.23	H,R,P
6f1c			Complement C1s	23	8.7	0.38981	2.02	H
1x5y			Fibronectin type-III domain of mouse myosin-binding protein C	28	14.3	0.42255	2.45	R
2cum			Fibronectin type III domain of human Tenascin-X	29	3.4	0.40251	2.8	R
		3f5n	Neuroserpin	41	12.2	0.36905	3.58	H
	3h6s		Clitocypin-5 cysteine proteinase	34	2.9	0.32861	3.45	H
		2yhf	C-type lectin domain family 5 member A	30	3.3	0.30137	3.42	H
	3bn3		Intercellular adhesion molecule 5	39	3.34	0.36204	3.34	H
	3sq9	3u8m	Acetylcholine-binding protein	38;38	13.2;10.5	0.35462;0.38651	3.69;3.14	R

alignment in complex with tissue factor (PDB: 1dan) showed potential for mimicry (Supplementary Fig. 2C) [51]. Interestingly, tissue factor expression has been shown to be up-regulated in sev-

ere SARS-CoV-2 infections, although there are several plausible theories [52,53]. The cystine disulfide loops of SARS-CoV and SARS-CoV-2 were found to resemble the EGF-like domains of

Table 2
Structural alignment values and data for exogenous hits.

Representative PDB			Mimic UniProt Name	Species	Alignment Length	Alignment Sequence Identity (%)	TM-score	RMSD	Tools
MERS-CoV	SARS-CoV	SARS-CoV-2							
1cej	2mgp	1n1i	Merozoite surface protein 1	Plasmodium knowlesi	22	4.5	0.25874	3.65	R
		2mgp	Merozoite surface protein 1	Plasmodium yoelii	25;24	4;4.2	0.25287;0.27536	2.98;3.33	R
		1ob1	Merozoite surface protein 1	Plasmodium falciparum	31;23	6.5;0	0.30361;0.24401	3.29;3.09	R
		1b9w	Merozoite surface protein 1	Plasmodium cynomolgi	24	4.2	0.26388	3.48	R
		2npr	Merozoite surface protein 1	Plasmodium vivax	24	4.2	0.24929	3.62	R
	1hn6	2j5l	Apical membrane antigen 1	Plasmodium falciparum	40;39	0;7.7	0.25857;0.26931	4.62;4.43	R,P
		1hky	Micronemal protein MIC4, related	Eimeria tenella (Coccidian parasite)	31;33	9.7;3	0.25771;0.31278	3.53;3.19	P
	4yiz	4yiz	Rhoptry neck protein 2, putative	Eimeria tenella (Coccidian parasite)	18;24;21	11.1;4.2;0	0.22386;0.28855;0.21409	2.79;3.29;3.36	P
	2j4w		Apical membrane antigen 1	Plasmodium vivax	26	7.7	0.29687	2.84	R,P
	2bbx		Thrombospondin-related anonymous protein	Plasmodium falciparum	25	4.8;13.5	0.30614	2.83	P
3sri			Rhoptry neck protein 2	Plasmodium falciparum (isolate 3D7)	17	5.9	0.28299	2.39	P
5wa2			Surface antigen 3	Toxoplasma gondii	36	8.3	0.50524	3.02	T
	4xvj	4g6a	E2 envelope glycoprotein	Hepatitis C virus	10;12	10;2.9	0.45252;0.37438	1.73;3.82	T,P
3kas			Transferrin receptor protein 1 / glycoprotein polyprotein GP complex	Machupo virus	38	10.5	0.51761	2.52	T
5f7l			Adhesin binding fucosylated histo-blood group antigen	Helicobacter pylori	37	2.7	0.53448	2.77	T
		2a2v	Kappa-theraphotoxin-Cg1a 1 (Jingzhaotoxin-XI)	Chilobrachys guangxiensis	18	0	0.23611	2.46	R
	2kni	2kni	Psalmotoxin-1	Psalmopoeus cambridgei	21;22	14.3;13.6	0.25651;0.25962	2.89;2.79	R,P
		1oma	Omega-agatoxin-Aa4b	Agelenopsis aperta	25	4	0.24137	3.07	R
	1qdp	1g1p	Delta-conotoxin EVIA	Conus ermineus	17	5.9	0.22626	3.01	R
		1qdp	Delta-hexatoxin-Ar1a (robustoxin)	Atrax robustus	26;23	7.7;0	0.25966;0.22613	3.76;3.7	R
		1la4	Kappa-theraphotoxin-Scg1a	Stromatopelma calceatum	22	13.6	0.22418	3.76	R
		2jtb	Hainantoxin-III 1	Haplopelma griseipes	18	5.6	0.22565	2.57	R
		1i26	Toxin Ptu1	Peirates turpis	21	9.5	0.25106	3.07	R
		1eit	Mu-agatoxin-Aa1a	Agelenopsis aperta	21	0	0.26268	2.73	R
		2mpq	Mu-theraphotoxin-Hd1a	Cyriopagopus doriae	18	5.6	0.27363	2.73	R
	1abt	2qc1	Alpha-bungarotoxin	Bungarus multicinctus	31;28	12.9;10.7	0.27217;0.33380	3.83;3.54	R
	4lft	4lft	Alpha-elapitoxin-Dpp2a	Dendroaspis polylepsis	26	0	0.25472;0.32330	3.38;2.63	R
		2jqp	Weak toxin 1	Bungarus candidus	26;25	7.7;4	0.30834	3.36	R
		2nbt	Kappa-bungarotoxin	Bungarus multicinctus	28	3.6	0.27782	3.74	R
		2ctx	Long neurotoxin 3	Naja naja	28	3.6	0.32545	3.61	R
	1lxh		Long neurotoxin 1	Naja kaouthia	23	0	0.27252	2.85	R
	1txa		Long neurotoxin 2	Ophiophagus hannah	28	3.6	0.29226	3.28	R
	1c6w		Maurocalcin	Scorpio palmatus	18	0	0.27771	2.6	R
4om5			Cytotoxin 4	Naja atra	31	9.7	0.40943	2.71	R
4om4			Cytotoxin 2	Naja atra	29	10.3	0.40317	2.31	R

(continued on next page)

Table 2 (continued)

Representative PDB			Mimic UniProt Name	Species	Alignment Length	Alignment Sequence Identity (%)	TM-score	RMSD	Tools
MERS-CoV	SARS-CoV	SARS-CoV-2							
1onj			Cobrotoxin-b	Naja atra	30	10	0.38853	2.56	R
1era			Erabutoxin b	Laticauda semifasciata	34	17.6	0.3819	3.14	R
5ebx			Erabutoxin a	Laticauda semifasciata	36	16.7	0.38112	3.49	R
2mj4			Short neurotoxin 1	Naja oxiana	35	17.1	0.36641	3.33	R
1cod			Cobrotoxin homolog	Naja atra	34	11.8	0.35821	3.29	R
3hh7			Haditoxin	Ophiophagus hannah	34	8.8	0.37387	3.42	R
1jgk			Candoxin	Bungarus candidus	33	6.1	0.37726	3.47	R
1je9			Cobrotoxin-c	Naja kaouthia	31	16.1	0.34998	3.02	R

coagulation factors X and IX and fibrillin, which are known to bind calcium [54,55]. However, there is no evidence for calcium binding to the RBMs.

All three RBMs were found to mimic the EGF-like domain of thrombomodulin, specifically in the region that binds thrombin (PDB: 1dx5), by RUPEE and PDBeFold, while the SARS-CoV-2 similarity was also detected by HMI-PRED [56]. Studies have shown that both thrombin and thrombomodulin blood concentrations are correlated with SARS-CoV-2 infection severity [57,58]. The verification by three tools and relevance to the literature led us to explore the potential mimicking of thrombomodulin binding to thrombin by the SARS-CoV-2 RBM using protein–protein docking (Fig. 3A). Calculation of the interaction energies revealed that the reference docking and experimental controls showed similar affinities of -6.12 and -6.88 kJ/mol, and the SARS-CoV-2 RBM bound at a slightly lower affinity of -1.96 kJ/mol. The similarity to thrombomodulin might help explain the prothrombotic coagulopathy presented in SARS-CoV-2 infections [59].

The central beta-strands of the SARS-CoV-2 RBM were found to be structurally similar to the transforming growth factor alpha, epiregulin, and epigen EGF-like domains using RUPEE; however, alignment of the RBM with the proteins in complex with the epidermal growth factor receptor (EGFR) (PDBs: 1mox, 5wb7, 5wb8) showed that the RBM was just out of ligand-binding range (Supplementary Fig. 2E) [60,61]. There is no evidence for interaction of the SARS-CoV-2 RBM with the extracellular domain of EGFR; however, the alignments were included for potential off-target effects related to the EGF-like domains.

Structural mimicry of chemokine and cytokine signaling has been reported for several viruses [1]. Viral proteins can mimic the chemokine, as in the case of HIV gp120 and CCL5, or they can mimic the receptor and bind directly to the cytokine (inhibiting its function), such as the vaccinia virus B15R protein that mimics the IL-1B receptor and binds to IL-1B [62,63].

Several cell signaling ligands and receptors were found similar to the coronavirus RBMs. The SARS-CoV-2 and SARS-CoV RBMs were both found to mimic IL-8 like chemokines, fibroblast growth factor 1, C-C motif chemokines 3, interleukin-18, and ephrins; they also individually mimic BMP2 and von Willebrand factor, respectively. The MERS-CoV RBM structurally resembled C-X-C motif chemokines 2 and 4 and growth/differentiation factor 5. The alignments of the RBMs with IL-8 like chemokines, C-C motif chemokines 2 (CCL2), 3, and 4, and IL-18 in complex with their respective receptors shows only partial alignment with the ligand-binding regions (Supplementary Fig. 2F). Interestingly, however, expression levels of these cytokines have all been shown as correlating with SARS-CoV-2 infection, although other explanations have been proposed [64–67]. For example, IL-33

release by damaged lower respiratory cells during SARS-CoV-2 has been demonstrated to trigger inflammation, increasing CCL2 and CCL3 expression [68].

Fibroblast growth factor 1 (FGF1) was shown to be similar to the SARS-CoV-2 and SARS-CoV RBMs using PDBeFold and HMI-PRED. A transcriptomic profiling revealed that FGF1 was upregulated in coronavirus infections [69]. Thus, to look more closely at potential interference, we docked the SARS-CoV RBM with the fibroblast growth factor receptor 2 (FGFR2) (PDB: 3OJ2), which was predicted by HMI-PRED (Fig. 3B) [70]. The RBM-FGFR2 docking analysis predicted a potentially favourable affinity of -2.49 kJ/mol, although not as high as the experimental complex (-9.78 kJ/mol) (Fig. 3D). The FGF1 signaling pathway may, thus, be modulated by the coronavirus spike RBMs.

HMI-PRED predicted that the SARS-CoV-2 RBM mimics ephrin-A5 and ephrin-B2 binding to the ephrin type 4a receptor (EPHA4), and SARS-CoV mimics ephrin-A5 binding to the ephrin receptor type 3a. EPHA4 is unique among known class A ephrin receptors in that it binds both ephrin a and b ligands [71,72]. The structural similarity of two ligands for the same receptor for the SARS-CoV-2 RBM motivated further testing with protein–protein docking with EPHA4. Although there is no evidence for ephrin receptor involvement in coronavirus infections, other viral surface proteins have been shown to utilize ephrin receptors for cell entry, such as the rhesus r virus [73]. The docking revealed similar affinities between the SARS-CoV-2 RBM-EPHA4, ephrin-A5-EPHA4, and experimental (PDB: 4m4r) complexes: -1.41 , -3.32 , -0.45 kJ/mol, respectively (Fig. 3A).

The platelet glycoprotein Ib (GP-Ib) binding domain of von Willebrand factor (VWF) was found to be similar to SARS-CoV by HMI-PRED. VWF-GP-Ib interaction has been shown as critical in modulating thrombosis and inflammation [74]. Although there is no literature on VWF and SARS-CoV infection, blood concentration levels of VWF have been shown as correlated with SARS-CoV-2 infection severity, which may indicate potential pathway interference [75].

SARS-CoV-2 was predicted to mimic bone morphogenetic protein 2 binding to activin receptor type-2B and MERS-CoV to mimic growth/differentiation factor 5 (GDF5) binding to bone morphogenetic protein receptor type-1A (BMPRT1A) by HMI-PRED; however, no experimental evidence is available for either case. To explore the potential involvement of the MERS-CoV RBM in cell signaling, we docked the MERS-CoV RBM to the BMPRT1A in the GDF5-binding region (PDB: 3qb4) [76]. The docking of the RBM and GDF5 displayed similar affinities to BMPRT1A with -6.73 and -6.99 kJ/mol, respectively, while the experimental complex bound with -10.58 kJ/mol.

RUPEE detected structural resemblance between the SARS-CoV-2 RBM and IL-6 receptor alpha and beta chains, both of which show

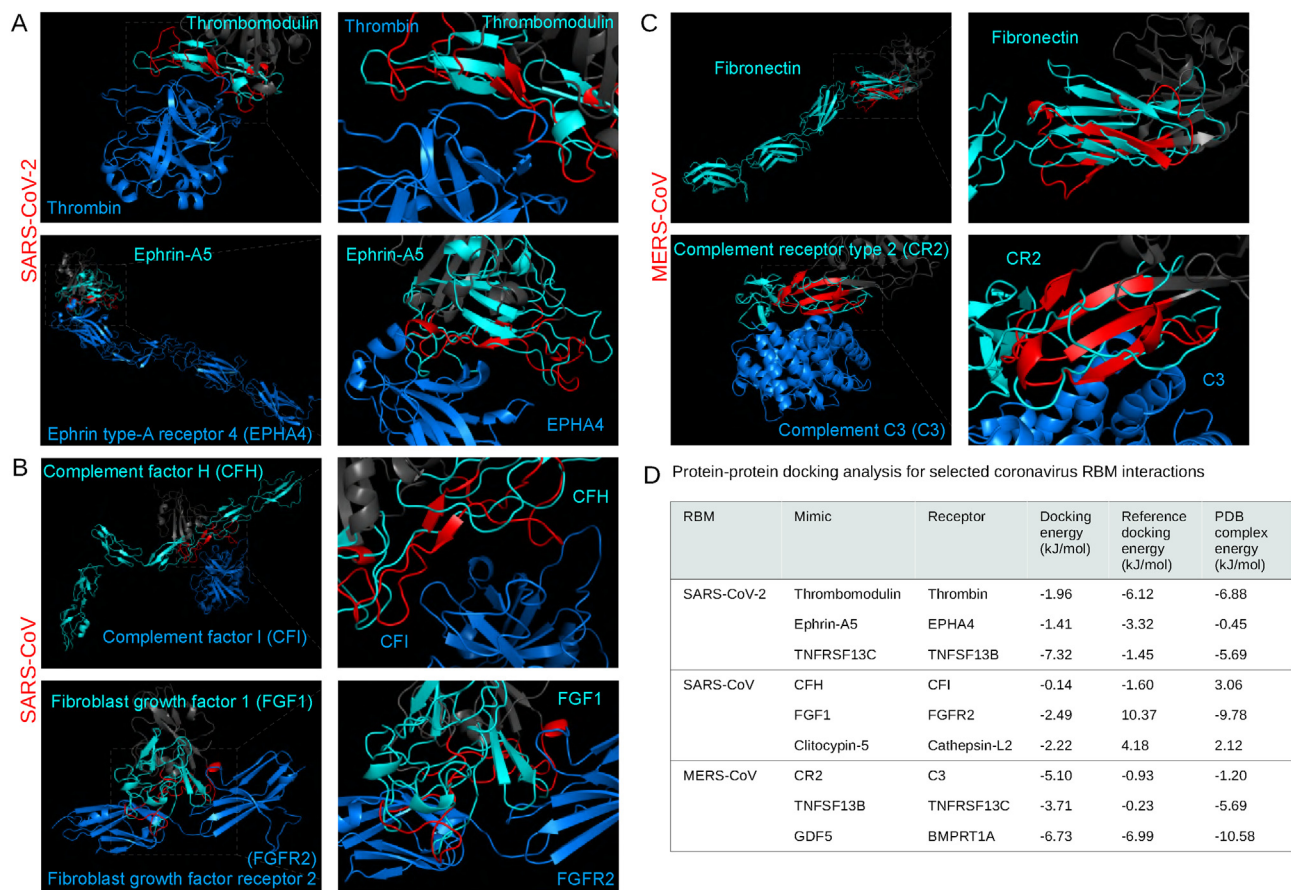


Fig. 3. Analysis of endogenous structural alignments. Protein-protein docking was performed using ClusPro PIPER to test the potential interactions between the coronavirus RBMs and potential interaction partners. The following alignments are shown: between the SARS-CoV-2 RBM and thrombin (A, top) and ephrin type-A receptor 4 (A, bottom), the SARS-CoV RBM and complement factor I (B, top) and fibroblast growth factor receptor 2 (B, bottom), and the MERS-CoV RBM with complement C3 (C, bottom). The MERS-CoV RBM aligned to fibronectin type III domain (C, top). RBMs are labelled red, the remainder of the RBMs are dark gray, mimicked proteins are cyan, and potential interaction partners are marine blue. Interaction energy scores predicted using FoldX on docked and experimental complexes (D). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

mimicry of the IL-6 binding sites (Supplementary Fig. 2G). IL-6 has been reported as an overexpressed cytokine in SARS-CoV-2 infections, which can lead to induction of a hyper-innate inflammatory response [77,78]. Mimicry of the IL-6 receptors by the RBM could result in binding and, thus, interference of IL-6 related interactions. However, several alternative theories have been proposed to explain the increases in IL-6 during severe infection; for example, the SARS-CoV nucleocapsid protein has been shown to activate IL-6 expression [79]. HMI-PRED additionally predicted MERS-CoV RBM mimicry of the binding of the T cell receptor beta chain to the major histocompatibility complex class I-related gene protein and interferon lambda receptor 1 binding to the beta subunit of the interleukin-10 receptor, both of which could have implications in immunosurveillance and inflammatory pathways [80,81].

Different tumor necrosis factor-related ligands and receptors were found to be structurally analogous to the MERS-CoV RBM and SARS-CoV-2 RBM, respectively. Tumor necrosis factor receptor superfamily (TNFRSF) 1A, 4, 13C, and 14 were returned for the cysteine disulfide loop for the SARS-CoV-2 RBM by RUPEE, while the tumor necrosis factor ligand superfamily (TNFSF) 13B and 14 were found to resemble the MERS-CoV RBM by RUPEE and HMI-PRED. Similarity of the SARS-CoV-2 RBM to TNFRSF 13C was also found by PDBFold. These signaling pathways normally promote B-cell and the T-cell survival and maturation [82,83]. The structural similarity of this family of ligands and receptors to the SARS-CoV-2 and MERS-CoV RBMs led us to further inspect the

interactions with protein-protein docking: mimicry of SARS-CoV-2 to TNFRSF 13C and MERS-CoV to TNFSF 13B. Thus, we simulated the binding of the SARS-CoV-2 RBM to TNFSF 13B and MERS-CoV to TNFRSF 13C (PDB: 3v56) [84]. Both cases revealed that the RBM is predicted to dock at a higher affinity than the natural ligand (Fig. 3D).

The complement system comprises a series of protein cascades that form an integral part of the innate immune response to viruses [85]. Viruses are generally susceptible to the complement system; however, viral proteins can utilize complement proteins through molecular mimicry in a variety of ways, such as using complement receptors for viral entry or evading detection by the immune system [86]. Infections from all three highly pathogenic coronaviruses have been reported to activate the complement system, enhancing pathogenicity, although the exact mechanisms remain unclear [87]. The spike protein of SARS-CoV-2 has been shown to localize near C4d and C5b-9 in lung vasculature, and mutations in several complement activation proteins, such as complement factors H, I, and III, have been found to correlate with infection severity [88,89]. The structural similarity screen yielded three motifs from the complement system that potentially mimic RBMs: complement factor I (CFI) binding domain of CFH for all three RBMs and both the complement C3d binding domain of complement receptor 2 (CR2) and the complement C1r binding domain of complement C1s for the MERS-CoV RBM. Interestingly, CFH and the SARS-CoV-2 spike protein have been proposed to compete for heparan

sulfate binding [90]. The SARS-CoV RBM, however, was predicted to be similar to CFH by RUPEE, PDBeFold, and HMI-PRED; thus, we docked the SARS-CoV RBM to CFI (PDB: 5o32) and found that the natural ligand was predicted to bind at a slightly higher affinity than the SARS-CoV RBM: -1.60 vs. -0.14 kJ/mol, respectively [91]. The C3d-binding domain of CR2 for the MERS-CoV RBM was also identified by RUPEE, PDBeFold, and HMI-PRED and was, thus, explored with docking of the MERS-CoV RBM to C3 (PDB: 3oed) (Fig. 3C) [92]. The MERS-CoV RBM was predicted to bind at a higher affinity than both the control docking and experimental complexes: -5.10 vs. -0.93 and -1.20 kJ/mol, respectively (Fig. 3D). Additionally, HMI-PRED found that the MERS-CoV RBM also mimics the complement C1r binding site of complement C1s. Additional experimental efforts are needed to validate the relationship between coronavirus spike proteins and the complement activation pathway.

Other endogenous hits included several unrelated proteins, such as protease inhibitors and serotransferrin. The MERS-CoV RBM resembled the fibronectin type III (FNIII) domains of mouse myosin-binding protein C and tenascin-X using RUPEE. Although myosin-binding protein C is intracellular, FNIII domains are found across the domains of life and function in diverse ways, from cell adhesion to cell signaling [93]. Drayman *et al.* found that the West Nile virus envelope glycoprotein E resembles the structural architecture of the FN10 domain of fibronectin, which is a natural ligand for integrin $\alpha\beta3$. Thus, we checked and found that the MERS-CoV RBM shares structural properties with other FNIII domains, such as those from fibronectin and neural cell adhesion molecule 1 (PDBs: 2haz and 1fnf, respectively) (Fig. 3C) [94,95]. The MERS-CoV RBM was also found to mimic part of the jagged-2, DLL1, and DLL4 proteins; however, the alignment was largely out of ligand-binding range when compared to jagged-1 in complex with NOTCH1 (PDB: 5uk5) – although the alignment may be relevant in other scenarios (Supplementary Fig. 2E) [96]. Protease inhibitors included neuroserpin for the SARS-CoV-2 RBM and cliticypin-5 cysteine protease for the SARS-CoV RBM. The alignment of the SARS-CoV with cliticypin-5 cysteine protease showed potential binding to cathepsin L2 (PDB: 3h6s) [97]. The role of cathepsins in coronavirus cell entry has been described as helping process the spike protein for viral and host membrane fusion [98]. To investigate the potential for additional interactions between coronavirus RBMs and cathepsins, we performed protein–protein docking. The binding of the SARS-CoV RBM to cathepsin L2 was predicted to be more favourable than the docking and experimental controls (Fig. 3D). Experimental evidence is required to validate this interaction, however. Both the SARS-related RBMs resembled motifs of serotransferrin using TM-align, and, interestingly, the transferrin receptor protein 1 has been proposed as a potential cell entry receptor for SARS-CoV-2 [17]. However, the alignments were generally out of ligand-binding range (PDB: 1suv) (Supplementary Fig. 2H); since no binding mode was apparent, it was not considered for docking [99]. HMI-PRED predicted that the SARS-CoV-2 RBM mimics the dimerization domain of C-type lectin domain family 5 and that the SARS-CoV RBM mimics intercellular adhesion molecule 5 binding to integrin α -L [100,101]. Integrins have been proposed to bind to the SARS-CoV-2 spike protein, although that is due to a new RGD motif in the RBD – of note, the RGD motif is not included in the selected residues for this study's SARS-CoV-2 RBM since it does not interact with ACE2 in experimental models [16]. Because integrin binding has not been hypothesized outside of the discussion of the SARS-CoV-2 RGD motif, docking was not pursued. Both the SARS-CoV-2 and SARS-CoV RBMs mimicked the nicotine-binding domain of the nicotinic acetylcholine receptor by RUPEE, which may have implications in the 'nicotinic hypothesis' [102].

2.5. Exogenous

We classified the exogenous hits by the pathogen type. There were motifs from apicomplexan parasites, viruses, one bacterial protein, and snake and spider toxins found to resemble the coronavirus RBMs.

The EGF-like domains from merozoite surface protein 1 (MSP1) of several *Plasmodium* species were found to be structurally similar to all three RBMs using RUPEE. Compared to the other two, the SARS-CoV-2 RBM was found to be similar to the most *Plasmodium* species: *falciparum*, *yoelii*, *cynomolgi*, *knowlesi*, *vivax*. The SARS-CoV RBM returned *P. yoelii* MSP1 and the MERS-CoV RBM returned *P. falciparum* MSP1. A closer inspection at the *P. falciparum* MSP1 alignments revealed that two EGF-like domains on the same PDB structure (1ob1) were found to resemble the SARS-CoV-2 RBM (Fig. 4A) [103]. The PDB structure is originally modelling the antibody-binding epitope of the EGF-like domain of MSP1; however, the antibody epitope is located on a loop just outside of the EGF-like domain. Thus, antibody-binding to the SARS-CoV-2 RBM could not be verified, but the presence of two EGF-like domains near an epitope may motivate experimental testing. The *P. falciparum* apical membrane antigen 1 epitope (PDB: 2j5l) was also found to resemble the SARS-CoV-2 and SARS-CoV RBMs, although, as in the case of MSP1, both RBMs aligned to a region outside of the antibody-interacting residues (Fig. 4B) [104]. These EGF-like domains from *Plasmodium* parasites may provide structural epitope scaffolding for cross-reactivity against the coronavirus spike RBMs [105]. Recent studies have pointed to a potential protective effect of *P. falciparum* infections against SARS-CoV-2 infection, although direct experimental evidence is yet to be established [106–109]. The MERS-CoV RBM was also found to resemble the rhoptry neck protein 2 and thrombospondin-related anonymous protein from *P. falciparum*. The surface antigen 3 of *Toxoplasma gondii* was found to be similar to the MERS-CoV RBM (Fig. 4C). Although there are no data on MERS-CoV and *T. gondii* co-infections, SARS-CoV-2 has been shown to have negative covariation with toxoplasmosis, which may indicate a protective effect from *T. gondii* [110].

The coronavirus RBMs were found to structurally mimic several motifs on the HIV and Influenza spike proteins; however, they were found either facing inwards or buried inside the mimicked protein and were, therefore, discarded. PDBeFold and TM-align indicated that the SARS-CoV-2 and SARS-CoV RBMs structurally mimic several hepatitis C virus (HCV) antibody epitopes. The SARS-CoV-2 and SARS-CoV RBMs were found to be similar to 10 and 6 PDB HCV E2 protein epitopes structures, respectively (Supplementary Table 2). The HCV E2 protein is implicated in host entry, which has been explored as an inhibitory target with neutralizing antibodies [111,112]. A closer inspection of the mapping of the epitopes to the SARS-CoV-2 RBM show that they are distributed across the RBM (Fig. 4D). Some studies have suggested that HCV may be negatively correlated with SARS-CoV-2 infection [113–115]. Since several of the epitopes were aligned in ways that were accessible to antibodies in the original PBD, we selected three epitopes, one at each region of the RBM, and docked the respective antibody to the SARS-CoV-2 RBM using the ClusPro PIPER 'antibody' mode (Fig. 4D). As shown in Fig. 4E, the RBM-antibody docking results were compared to docking and experimental controls – the antibodies bound in a similar way to docking controls in all three cases, while the experimental complexes were predicted to bind more tightly. These structural similarities may take part in potential cross-reactivity between HCV and coronavirus infections. Of note, two recently proposed cell entry receptors for the SARS-CoV-2 spike protein, ASGR1 and APOA4, have been shown as potentially implicated in mediating HCV viral entry [116]. In an

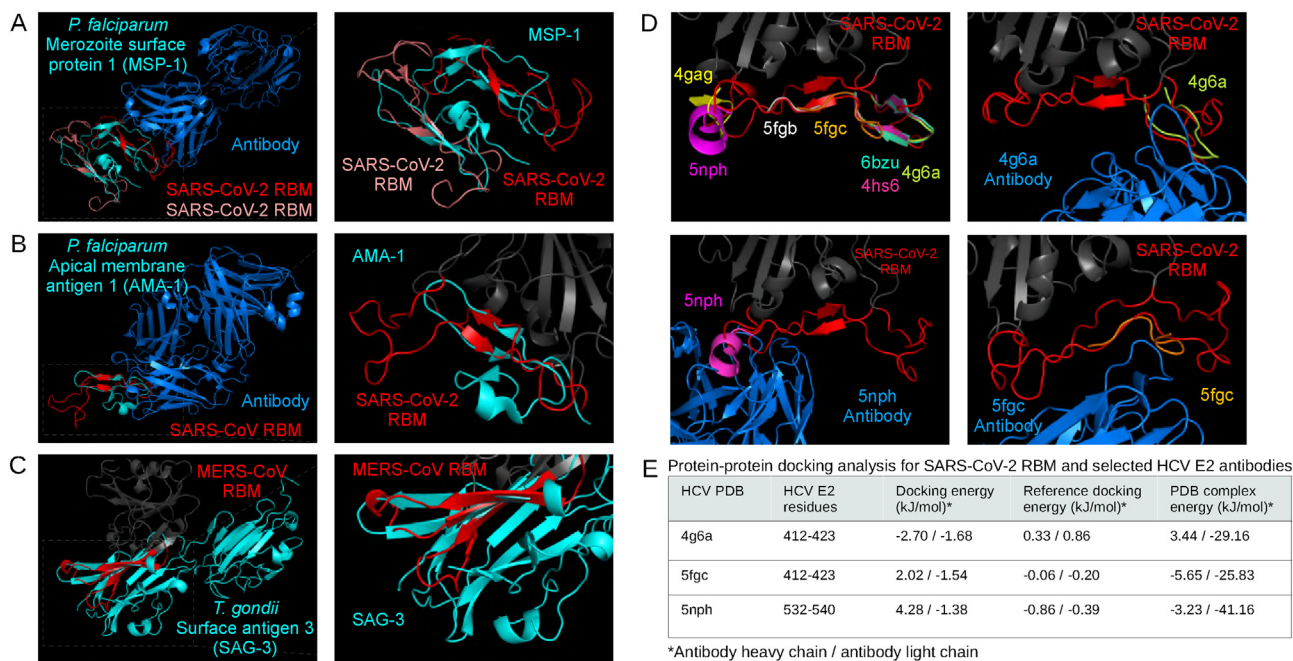


Fig. 4. Analysis of exogenous structural alignments. *Plasmodium falciparum* merozoite surface protein 1 (A) and apical membrane antigen 1 (B) structurally aligned with the SARS-CoV-2 RBM and SARS-CoV RBM, respectively. The *Toxoplasma gondii* surface antigen 3 aligned with the MERS-CoV RBM (C). RBMs are labelled red, mimicked proteins are cyan, and potential interaction partners are marine blue (A, B, C). Hepatitis C virus epitopes structurally aligned to the SARS-CoV-2 RBM, and the respective antibody structures from PDBs 5fgc, 5nph, 4g6a docked to the RBM (D) using ClusPro PIPER “antibody” mode. Interaction energy scores predicted using FoldX on docked and experimental complexes (E). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

interesting case, the MERS-CoV RBM was found to structurally mimic both the Machupo virus glycoprotein polyprotein GP complex RBM (TM-score: 0.47) and its receptor, transferrin receptor protein 1 (0.52) (PDB: 3kas) using TM-align, although the transferrin receptor scores slightly higher (Supplementary Fig. 3A) [117].

Only one bacterial protein was selected in the structural similarity screen. The adhesin-binding fucosylated histo-blood group antigen of *Helicobacter pylori* was found to be similar to the MERS-CoV RBM by TM-align. The structure (PDB: 5f7l) shows binding of the bacterial protein to a nanobody; however, the RBM alignment is just outside of the nanobody binding site (Supplementary Fig. 3B) [118]. No studies have detailed any connections between MERS-CoV and *H. pylori*.

Motifs from snake, spider, and cone snail toxins were found to be similar to all three RBMs using PDBFold and RUPEE. The SARS-CoV-2 and SARS-CoV RBMs shared similarity to four toxins, and the MERS-CoV RBM only returned unique proteins. The two SARS-related viruses mimicked three-finger bungarotoxins and inhibitor cystine-knot toxins, such as psalmotoxin-1, while MERS-CoV RBM resembled other three-finger toxins, like cytotoxin 4 (Supplementary Fig. 3C, D) [119–123]. In total, these toxins may bind to several receptors involved in nociception, e.g. ASIC1 and Nav1.7, which may be relevant to the taste and pain perception changes experienced during SARS-CoV-2 infection [124]. Importantly, and perhaps confoundingly, a recent study found no changes in depolarization for Nav1.7 and Cav2.2 upon exposure to the SARS-CoV-2 RBD [125]. Thus, further experimental work is necessary to validate these interactions.

3. Methods

3.1. Spike receptor-binding motif model generation and characterization

Amino acid sequences of the SARS-CoV-2 (NCBI code: NC_045512), SARS-CoV (NC_004718), and MERS-CoV (NC_038294) spike proteins were extracted as FASTA files from

the NCBI Viral Genomes Resource [133]. Each amino acid sequence corresponds to one of three identical protomers of the full homo-oligomeric spike trimer. Due to the high number of available experimentally-resolved structures for each spike protein, representative models were generated using ProtCHOIR – a recently developed bioinformatic tool to automate 3D homology modelling of homo-oligomers [134]. ProtCHOIR builds homo-oligomeric assemblies by searching for homolog templates on a locally created homo-oligomeric protein database using PSI-BLAST, performing a series of structural analyses on the input protomer structure or sequence using Molprobit, PISA, and GESAMT (all three tools as part of the CCP4 Molecular Graphics package), and comparative homology modelling using MODELLER (version 9.24) with molecular dynamics-level optimization and refinement [135–140]. Trimerization was detected for all three coronavirus spike proteins.

The residues for the receptor-binding domains (RBD) and RBMs of each spike model were manually selected based on experimental structures with primary receptors (residues defined in Supplementary Table 1) and made into sub-structures during manual inspection of full-length models on PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). Global amino acid sequence alignment of RBDs was performed with EMBOSS Needle [141]. The full-length SARS-CoV-2 spike protein modelled with the lipid bilayer displayed in Fig. 1 was retrieved from the SARS-CoV-2 3D database [142]. No records exist of N-linked or O-linked glycosylation motifs near the three RBMs, which was supported by NetNGlyc 1.0 and NetOGlyc 4.0 predictions [143,144]. To determine the flexibility of each residue in the RBDs, we used CABS-flex 2.0, a web server that offers fast simulations and resulting data of protein structure flexibility [145]. Default values were used and residue flexibility was reported as root mean squared flexibility (RMSF) [146].

3.2. Structure similarity screen

Several web servers and stand-alone tools have become available to perform pairwise or multiple sequence-independent

protein structure alignments, such as DALI, FATCAT, iSARST, MADOKA, PDBFold, TM-align, and RUPEE [147–153]. After testing each tool, the PDBFold web server, RUPEE web server, and a locally-installed version of TM-align were selected due to the diversity of structural alignment methodologies, ease-of-use, data accessibility, and widespread-usage. A newly published web server for structural prediction of host-microbe interactions based on interface mimicry, HMI-PRED, was also included in the analysis.

Of note, mTM-align (the web server version of TM-align) was considered, but no non-spike proteins were shown – restricting the downstream analysis [154]. Thus, all 3D models in the PDB database clustered at 100% sequence identity were downloaded, and TM-align was run in a pairwise manner, using GNU parallel, between each RBM model and each chain of every downloaded PDB file (O. Tange (2018): GNU Parallel, March 2018, <https://doi.org/10.5281/zenodo.1146014>). TM-align works by, first, combining secondary structure similarity alignments, defined by DSSP (Define Secondary Structure of Proteins), and TM-score-based structural alignments [155,156]. A structure rotation matrix is applied to the alignments in order to maximize the TM-score, which was used to rank the alignments for each RBM.

The PDBFold web server utilizes SSM, a graph-matching algorithm that superimposes *PROMOTIF*-defined secondary structures and, subsequently, maps backbone carbon atoms of, first, matched and, second, unmatched secondary structures [151,157,158]. The hits are ranked by their Q-Score, which is calculated to achieve a lower root mean squared deviation (RMSD) and an increased number of aligned residues. Since the highest percentage (%) of secondary structure matches for the SARS-CoV RBM was found at 67% (while SARS-CoV-2 and MERS-CoV returned hits with 100% secondary structure matches), we set the PDBFold search parameters for 65% structural similarity at “highest precision” for each RBM.

The RUPEE web server performs structural similarity comparisons using a purely geometric approach: 1) a linear encoding of

the protein structure is defined to identify separable regions of permissible torsion angles for DSSP secondary structure assignments; 2) the encoding is converted into a bag of features; 3) a protein structure indexing method is established using min-hashing and locality sensitive hashing; 4) the top 8,000 matches are sorted based on adjusted Jaccard similarity scores; and 5) if running in “Top-Aligned” mode (used in this analysis), the alignments are re-scored using TM-align [153]. RUPEE allows the specific comparison of a query protein to the CATH, SCOP, PDB, and ECOD databases [159–162]. Additional settings are also offered, such as the “contains” (finding query protein inside database protein) and “contained in” options (small protein motif detection in query protein) – both of which were used in this analysis.

HMI-PRED combines TM-align with NACCESS to search through template host protein-protein complexes [163,164]. The structural alignment model of the ligand and putative receptor are refined with RosettaDock to quantify electrochemical complementarity, which is not included in a strict structural alignment screen [165]. Alignments for both the RBM and RBD of each of the three coronaviruses were collected from HMI-PRED.

Since the flexibility of the cystine disulfide loop on the SARS-CoV-2 and SARS-CoV RBMs may affect the global structure of the RBMs and, thus, search outcome, we used two additional models provided by CABS-flex 2.0 for both RBMs, making a total of three conformations for each SARS-CoV-2 and SARS-CoV. Results of the different conformations provided were pooled together for both SARS-related RBMs. The top-scoring alignments from all four tools for each RBM were matched with their corresponding PDB and, subsequently, UniProt accession code [166]. The UniProt accession codes were then compared across tools to identify shared top hits.

3.3. Protein-protein docking and interaction energy prediction

A local installation of ClusPro PIPER (version 1.1.5) was used for protein-protein docking [167–170]. Annotations informing

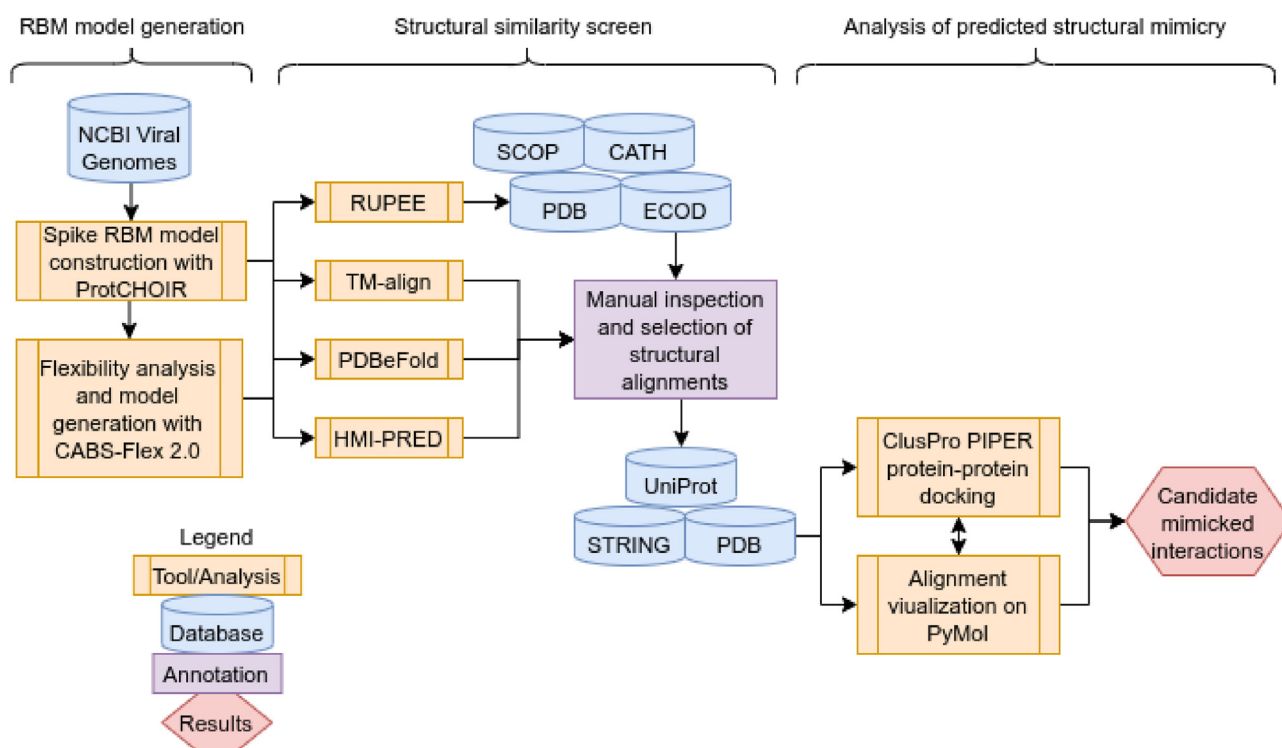


Fig. 5. Pipeline flow. A flow chart of the analyses performed in this study.

potential protein-protein interactions were obtained from the PDB, STRING, and UniProt databases [171]. The ClusPro PIPER “antibody” docking mode was used to dock RBDs with the hepatitis C antibody PDB structures (PDBs: 4g6a, 5fgc, 5nph), and the “others” mode was used for all non-antibody docking. In order to minimize non-biologically relevant binding during the docking runs, residues outside of the RBM on the spike RBDs and outside of the ligand-binding region on the predicted interaction partners were masked. The docked models were minimized using CHARMM22. To gain better insight into the binding strength of the potential RBD-receptor complex interactions, we used the FoldX (version 4.0) AnalyseComplex program, which predicts the interaction energy by finding the difference in stability between the individual unfolded molecules and the overall complex [172]. The original PDB ligands were also docked to the receptor in order to obtain a “Reference docking energy” when compared with the predicted RBD-receptor energy. The “PDB complex energy” was obtained from the original PDB containing the ligand and receptor to understand binding resolution of the experimental complex.

3.4. Data analysis and visualization

A full representation of the pipeline and tools used can be found in Fig. 5. Data were analyzed and plots were generated using R version 3.6.3 (2020-02-29). Protein structural alignments were visualized with PyMOL (version 1.8.4.0) [173]. Pdb-tools was used to manipulate and organize PDB files [174]. The graphical abstract was adapted from the “SARS-CoV-2 Spike Protein Conformations” template on BioRender. Fig. 5 was created on draw.io. Raw data and alignment models are made available at <https://github.com/tlb-lab>.

4. Conclusions

This study involved the structural bioinformatics characterization of potential molecular mimicry by highly pathogenic coronavirus spike protein RBMs. Using protein homology modelling, we built representative models of the spike RBMs and tested structural changes in the SARS-CoV-2 RBM induced by recently recorded mutations, which had little effect on overall RBM structure. Comparison of the RBMs revealed that the SARS-CoV and SARS-CoV-2 RBMs share higher structural homology than with MERS-CoV, which was underlined by the number of common returned proteins in the structural similarity screen using four structural alignment tools. The flexibility of the cystine disulfide loop in the SARS-related RBMs was found to permit large global changes in RBM structure; however, since most of the predicted mimicry was mapped to the RBM central beta-strands, which are quite rigid, the models of different conformations did not return significantly different proteins from the structural alignment tools. The structural alignment screen highlighted the similarity of the RBMs to evolutionarily unrelated human and non-human proteins. Further validation of the alignments with protein-protein docking revealed that all tested coronavirus RBM-endogenous protein interactions were predicted to be energetically favourable, confirming that the structural similarity screen may be useful in identifying potential molecular mimics.

The predicted endogenous mimicry comprised of proteins in cell signaling, adhesion, and complement pathways. Potential mimicry of several microbial antigenic proteins and exogenous toxins was also discovered. The EGF-like domains of both endogenous and exogenous proteins structurally resemble all three RBMs. Predicted mimicked endogenous interactions include the EGF-like domain of thrombomodulin binding to thrombin, NOTCH1 binding to DLL4, and coagulation factor VIIa binding to tissue factor. Inter-

ference in these pathways may partially explain coagulopathies in coronavirus infections [126]. Exogenous EGF-like domains of MSP1 from different *Plasmodium* species, on the other hand, may provide a structural epitope scaffold for cross-reactivity between coronavirus and *Plasmodium* infections [106]. Epitope similarity was further explored among the several antibody-bound hepatitis C virus E2 protein motifs that were structurally analogous to the SARS-related RBMs. Structural similarity to antigenic proteins from other microbes may confer cross-immunity and, thus, also potentially guide vaccine design [127]. Cell signaling pathway proteins, such as TNF-related and ephrin ligands, were also found as potential mimics of the coronavirus RBMs, which may lead to use of alternative co-receptors for viral entry or modulation of signaling cascades. Complement factor H was returned for all three RBMs and has also been implicated in coronavirus infections [90]. The mimicry of complement proteins is widespread among viruses, and the spike RBM may have secondary roles interfering in these pathways [128]. Many snake and spider toxins were also found similar to the coronavirus RBMs, which implies the potential usage of receptors involved in pain, muscle contraction, cell adhesion, and coagulation pathways [129–131]. The prediction of evolutionarily unrelated, yet structurally similar, potential protein mimics reveals that previously unidentified pathways could be altered by the spike RBMs. The structural variation between coronavirus RBMs and their resulting molecular mimics can possibly be connected to differences in tropism, infection severity, and immune system reactivity between coronaviruses.

Although experimental verification of the predicted interactions is required to take these results further, the findings presented in this study provide insight into the potential molecular mimicry utilized by highly pathogenic coronavirus RBMs. The data can be used to support inhibitory drug, peptide, and antibody design efforts in order to prevent viral cell entry and virulence mechanisms related to coronavirus RBMs [132]. Additional work is needed to better understand how coronaviruses co-opt host machinery to enhance fitness.

5. Permission note

Permission has been granted to use all text, illustrations, charts, tables, photographs, or other material from previously published sources.

CRediT authorship contribution statement

Christopher A. Beaudoin: Conceptualization, Methodology, Validation, Formal analysis, Investigation. **Arian R. Jamasb:** Methodology, Formal analysis, Investigation. **Ali F. Alsulami:** Formal analysis, Investigation, Resources. **Liviu Copoiu:** Validation, Formal analysis, Investigation. **Andries J. van Tonder:** Software, Resources. **Sharif Hala:** Software, Resources. **Bridget P. Bannerman:** Software, Resources. **Sherine E. Thomas:** Software, Resources. **Sundeeep Chaitanya Vedithi:** Software, Funding acquisition. **Pedro H.M. Torres:** Software, Validation, Supervision. **Tom L. Blundell:** Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Many thanks to the ClusPro PIPER team for letting our team download and install their software. A big thank you to the BioRender

team and Melissa Pappas for help with figure rendering. CAB, SCV, and TLB were supported by American Leprosy Missions, United States of America (Grant No: G88726). TLB thanks the Wellcome Trust for support through an Investigator Award (200814/Z/16/Z; 2016 - 2021). CAB is funded by Antibiotic Research UK (PHZJ/687). ARJ is funded by the Biotechnology and Biological Sciences Research Council (BBSRC) DTP studentship (BB/M011194/1). AFA is funded by the King Abdullah scholarship (Saudi Arabia Research Council). LC and TLB are being funded by Ipsen Bioinnovation (PHZJ/487 - RG83343). AvT was supported by a joint Biotechnology and Biological Sciences Research Council (BBSRC) research grant under grant number BB/N00468X/1. BPB is supported by The Wellcome Trust (107032AIA), The UK Cystic Fibrosis Trust (Innovation Hub grant 001) and NIHR Cambridge Biomedical Research Centre. SET is funded by the Cystic Fibrosis Trust (RG 70975) and Fondation Botnar (RG91317). PHMT is funded by the Brazilian National Council for Scientific and Technological Development (CNPq).

Author contributions

CAB performed analyses and wrote the paper; ARJ, AFA, and LC performed analyses and provided expertise on bioinformatics tools and statistical analysis; AvT, SH, BPB, and SET provided expertise on sequence and structural investigations; SCV helped acquire funding and provided expertise on the structural mutation analysis; PHMT and TLB supervised and provided overall direction for the study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2021.06.041>.

References

- [1] Alcamí A. Viral mimicry of cytokines, chemokines and their receptors. *Nat Rev Immunol* 2003;3(1):36–50. <https://doi.org/10.1038/nri980>.
- [2] Lasso G, Honig B, Shapira SD. A sweep of earth's virome reveals host-guided viral protein structural mimicry and points to determinants of human disease. *Cell Syst* 2021;12:82–91.e3. <https://doi.org/10.1016/j.cels.2020.09.006>.
- [3] Elde NC, Malik HS. The evolutionary conundrum of pathogen mimicry. *Nat Rev Microbiol* 2009;7(11):787–97. <https://doi.org/10.1038/nrmicro2222>.
- [4] Walls AC, Xiong X, Park Y-J, Tortorici MA, Snijder J, Quispe J, et al. Unexpected receptor functional mimicry elucidates activation of coronavirus fusion. *Cell* 2019;176(5):1026–1039.e15. <https://doi.org/10.1016/j.cell.2019.05.006>.
- [5] Tim Chew F, Ong SY, Hew CL. Severe acute respiratory syndrome coronavirus and viral mimicry. *Lancet* 2003;361(9374):2081. [https://doi.org/10.1016/S0140-6736\(03\)13652-5](https://doi.org/10.1016/S0140-6736(03)13652-5).
- [6] Martínez YA, Guo X, Portales-Pérez DP, Rivera G, Castañeda-Delgado JE, García-Pérez CA, et al. The analysis on the human protein domain targets and host-like interacting motifs for the MERS-CoV and SARS-CoV/CoV-2 infers the molecular mimicry of coronavirus. *PLoS ONE* 2021;16(2):e0246901. <https://doi.org/10.1371/journal.pone.0246901>.
- [7] Renu K, Prasanna PL, Valsala GA. Coronaviruses pathogenesis, comorbidities and multi-organ damage – A review. *Life Sci* 2020;255:117839. <https://doi.org/10.1016/j.lfs.2020.117839>.
- [8] Cheong KK. Systematic Review of Ocular Involvement of SARS-CoV-2 in Coronavirus Disease 2019. *Curr Ophthalmol Rep* 2020;8(4):185–94. <https://doi.org/10.1007/s40135-020-00257-7>.
- [9] Zheng B, Yuan M, Ma Q, Wang S, Tan Y, Xu Y, et al. Landscape of SARS-CoV-2 spike protein-interacting cells in human tissues. *Int Immunopharmacol* 2021;107:567. <https://doi.org/10.1016/j.intimp.2021.107567>.
- [10] Angileri F, Legare S, Marino Gammazza A, Conway de Macario E, JL Macario A, Cappello F. Molecular mimicry may explain multi-organ damage in COVID-19. *Autoimmun Rev* 2020;19(8):102591. <https://doi.org/10.1016/j.autrev.2020.102591>.
- [11] Reguera J, Mudgal G, Santiago C, Casasnovas JM. A structural view of coronavirus–receptor interactions. *Virus Res* 2014;194:3–15. <https://doi.org/10.1016/j.virusres.2014.10.005>.
- [12] Tortorici MA, Vesler D. Structural insights into coronavirus entry. *Adv Virus Res* 2019;105:93–116. <https://doi.org/10.1016/bs.aivir.2019.08.002>.
- [13] Li F, Goff SP. Receptor recognition mechanisms of coronaviruses: a decade of structural studies. *J Virol* 2015;89(4):1954–64. <https://doi.org/10.1128/JVI.02615-14>.
- [14] T. Ichimura Y. Mori P. Aschauer K.M. Padmanabha Das R.F. Padera A. Weins et al. KIM-1/TIM-1 is a Receptor for SARS-CoV-2 in Lung and Kidney MedRxiv 2020:2020.09.16.20190694. 10.1101/2020.09.16.20190694.
- [15] Yang C, Zhang Y, Chen H, Chen Y, Yang D, Shen Z, et al. Kidney injury molecule-1 is a potential receptor for SARS-CoV-2. *BioRxiv* 2020:2020.10.09.334052. <https://doi.org/10.1101/2020.10.09.334052>.
- [16] Sigrist CJ, Bridge A, Le Mercier P. A potential role for integrins in host cell entry by SARS-CoV-2. *Antiviral Res* 2020;177. <https://doi.org/10.1016/j.antiviral.2020.104759>.
- [17] Tang X, Yang M, Duan Z, Liao Z, Liu L, Cheng R, et al. Transferrin receptor is another receptor for SARS-CoV-2 entry. *BioRxiv* 2020:2020.10.23.350348. <https://doi.org/10.1101/2020.10.23.350348>.
- [18] Li W, Moore MJ, Vasileva N, Sui J, Wong SK, Berne MA, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 2003;426(6965):450–4. <https://doi.org/10.1038/nature02145>.
- [19] Wang N, Shi X, Jiang L, Zhang S, Wang D, Tong P, et al. Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res* 2013;23(8):986–93. <https://doi.org/10.1038/cr.2013.92>.
- [20] Gu Y, Cao J, Zhang X, Gao H, Wang Y, Wang J, et al. Interaction network of SARS-CoV-2 with host receptome through spike protein. *BioRxiv* 2020:2020.09.09.287508. <https://doi.org/10.1101/2020.09.09.287508>.
- [21] Patra T, Meyer K, Geerling L, Isbell TS, Hoft DF, Brien J, et al. SARS-CoV-2 spike protein promotes IL-6 trans-signaling by activation of angiotensin II receptor signaling in epithelial cells. *PLOS Pathog* 2020;16(12):e1009128. <https://doi.org/10.1371/journal.ppat.1009128>.
- [22] Buzhdygan TP, DeOre BJ, Baldwin-Leclair A, Bullock TA, McGary HM, Khan JA, et al. The SARS-CoV-2 spike protein alters barrier function in 2D static and 3D microfluidic in-vitro models of the human blood–brain barrier. *Neurobiol Dis* 2020;146:105131. <https://doi.org/10.1016/j.nbd.2020.105131>.
- [23] Rhea EM, Logsdon AF, Hansen KM, Williams LM, Reed MJ, Baumann KK, et al. The S1 protein of SARS-CoV-2 crosses the blood–brain barrier in mice. *Nat Neurosci* 2021;24(3):368–78. <https://doi.org/10.1038/s41593-020-00771-8>.
- [24] Ling R, Dai Y, Huang B, Huang W, Yu J, Lu X, et al. In silico design of antiviral peptides targeting the spike protein of SARS-CoV-2. *Peptides* 2020;130:170328. <https://doi.org/10.1016/j.peptides.2020.170328>.
- [25] Pandey P, Rane JS, Chatterjee A, Kumar A, Khan R, Prakash A, et al. Targeting SARS-CoV-2 spike protein of COVID-19 with naturally occurring phytochemicals: an in silico study for drug development. *J Biomol Struct Dyn* 2020:1–11. <https://doi.org/10.1080/07391102.2020.1796811>.
- [26] Hussain A, Hasan A, Nejadi Babadaei MM, Bloukh SH, Chowdhury MEH, Sharifi M, et al. Targeting SARS-CoV2 spike protein receptor binding domain by therapeutic antibodies. *Biomed Pharmacother* 2020;130:110559. <https://doi.org/10.1016/j.biopha.2020.110559>.
- [27] Kanduc D, Shoenfeld Y. Molecular mimicry between SARS-CoV-2 spike glycoprotein and mammalian proteomes: implications for the vaccine. *Immunol Res* 2020;68(5):310–3. <https://doi.org/10.1007/s12026-020-09152-6>.
- [28] Drayman N, Glick Y, Ben-nun-Shaul O, Zer H, Zlotnick A, Gerber D, et al. Pathogens use structural mimicry of native host ligands as a mechanism for host receptor engagement. *Cell Host Microbe* 2013;14(1):63–73. <https://doi.org/10.1016/j.chom.2013.05.005>.
- [29] Angileri F, Legaré S, Marino Gammazza A, Conway de Macario E, Macario AJL, Cappello F. Is molecular mimicry the culprit in the autoimmune haemolytic anaemia affecting patients with COVID-19? *Br J Haematol* 2020;190:e92–3. <https://doi.org/10.1111/bjh.16883>.
- [30] Huang AT, Garcia-Carreras B, Hitchings MDT, Yang B, Katzelnick LC, Rattigan SM, et al. A systematic review of antibody mediated immunity to coronaviruses: kinetics, correlates of protection, and association with severity. *Nat Commun* 2020;11(1). <https://doi.org/10.1038/s41467-020-18450-4>.
- [31] An H, Park J. Molecular Mimicry Map (3M) of SARS-CoV-2: Prediction of potentially immunopathogenic SARS-CoV-2 epitopes via a novel immunoinformatic approach. *BioRxiv* 2020:2020.11.12.344424. <https://doi.org/10.1101/2020.11.12.344424>.
- [32] Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. *Cell Host Microbe* 2020;27(4):671–680.e2. <https://doi.org/10.1016/j.chom.2020.03.002>.
- [33] Lin L, Ting S, Yufei H, Wendong L, Yubo F, Jing Z. Epitope-based peptide vaccines predicted against novel coronavirus disease caused by SARS-CoV-2. *Virus Res* 2020;288. <https://doi.org/10.1016/j.virusres.2020.198082>.
- [34] Lucchese G, Flöel A. Molecular mimicry between SARS-CoV-2 and respiratory pacemaker neurons. *Autoimmun Rev* 2020;19(7):102556. <https://doi.org/10.1016/j.autrev.2020.102556>.
- [35] Westall FC. Molecular mimicry or structural mimicry? *Mol Immunol* 2006;43(7):1062–4. <https://doi.org/10.1016/j.molimm.2005.06.039>.
- [36] Krishna SS, Grishin NV. Structurally analogous proteins do exist! *Structure* 2004;12:1125–7. <https://doi.org/10.1016/j.str.2004.06.004>.
- [37] Nwanochie E, Uversky VN. Structure determination by single-particle cryo-electron microscopy: only the sky (and Intrinsic Disorder) is the Limit. *Int J Mol Sci* 2019;20(17):4186. <https://doi.org/10.3390/ijms20174186>.
- [38] Wang L, Xiang Ye. Spike glycoprotein-mediated entry of SARS coronaviruses. *Viruses* 2020;12(11):1289. <https://doi.org/10.3390/v12111289>.
- [39] Nelson G, Buzko O, Spilman P, Niazji K, Rabizadeh S, Soon-Shiong P. Molecular dynamic simulation reveals E484K mutation enhances spike RBD-ACE2 affinity and the combination of E484K, K417N and N501Y mutations (501Y.V2 variant) induces conformational change greater than N501Y mutant alone.

- potentially resulting in an escap. *BioRxiv* 2021;2021.01.13.426558. <https://doi.org/10.1101/2021.01.13.426558>.
- [40] Saputri DS, Li S, van Eerden FJ, Rozewicki J, Xu Z, Ismanto HS, et al. Flexible, functional, and familiar: characteristics of SARS-CoV-2 spike protein evolution. *Front Microbiol* 2020;11. <https://doi.org/10.3389/fmicb.2020.02112>
 - [41] Spinello A, Saltalamacchia A, Magistrato A. Is the rigidity of SARS-CoV-2 spike receptor-binding motif the hallmark for its enhanced infectivity? insights from all-atom simulations. *J Phys Chem Lett* 2020;11(12):4785–90. <https://doi.org/10.1021/acs.jpclett.0c01148>
 - [42] Versteeg GA, van de Nes PS, Bredenbeek PJ, Spaan WJM. The coronavirus spike protein induces endoplasmic reticulum stress and upregulation of intracellular chemokine mRNA concentrations. *J Virol* 2007;81(20):10981–90. <https://doi.org/10.1128/JVI.01033-07>
 - [43] Doolittle JM, Gomez SM. Structural similarity-based predictions of protein interactions between HIV-1 and Homo sapiens. *Virol J* 2010;7:82. <https://doi.org/10.1186/1743-422X-7-82>
 - [44] Engel J. EGF-like domains in extracellular matrix proteins: Localized signals for growth and differentiation? *FEBS Lett* 1989;251:1–7. [https://doi.org/10.1016/0014-5793\(89\)81417-6](https://doi.org/10.1016/0014-5793(89)81417-6)
 - [45] Tombling BJ, Wang CK, Craik DJ. EGF-like and other disulfide-rich microdomains as therapeutic scaffolds. *Angew Chemie Int Ed* 2020;59:11218–32. <https://doi.org/10.1002/anie.201913809>
 - [46] D'Alonzo D, De Fenza M, Pavone V. COVID-19 and pneumonia: a role for the uPA/uPAR system. *Drug Discov Today* 2020;25:1528–34. <https://doi.org/10.1016/j.drudis.2020.06.013>
 - [47] Rovina N, Akinosoglou K, Eugen-Olsen J, Hayek S, Reiser J, Giamarellos-Bourboulis EJ. Soluble urokinase plasminogen activator receptor (suPAR) as an early predictor of severe respiratory failure in patients with COVID-19 pneumonia. *Crit Care* 2020;24:187. <https://doi.org/10.1186/s13054-020-02897-4>
 - [48] Shang Y, Smith S, Hu X. Role of Notch signaling in regulating innate immunity and inflammation in health and disease. *Protein Cell* 2016;7(3):159–74. <https://doi.org/10.1007/s13238-016-0250-0>
 - [49] Luca VC, Jude KM, Pierce NW, Nachury M V, Fischer S, Garcia KC. Structural basis for Notch1 engagement of Delta-like 4. *Science* (80–) 2015;347:847 LP – 853. <https://doi.org/10.1126/science.1261093>
 - [50] Rizzo P, Vieceli Dalla Sega F, Fortini F, Marracino L, Rapezzi C, Ferrari R. COVID-19 in the heart and the lungs: could we “Notch” the inflammatory storm? *Basic Res Cardiol* 2020;115:1–8. <https://doi.org/10.1007/s00395-020-0791-5>
 - [51] Banner DW, D'Arcy A, Chène C, Winkler FK, Guha A, Konigsberg WH, et al. The crystal structure of the complex of blood coagulation factor VIII with soluble tissue factor. *Nature* 1996;380(6569):41–6. <https://doi.org/10.1038/380041a0>
 - [52] Eslamifard Z, Behzadifard M, Soleimani M, Behzadifard S. Coagulation abnormalities in SARS-CoV-2 infection: overexpression tissue factor. *Thromb J* 2020;18:38. <https://doi.org/10.1186/s12959-020-00250-x>
 - [53] Bautista-Vargas M, Bonilla-Abadía F, Cañas CA. Potential role for tissue factor in the pathogenesis of hypercoagulability associated with in COVID-19. *J Thromb Thrombolysis* 2020;50(3):479–83. <https://doi.org/10.1007/s11239-020-02172-x>
 - [54] Handford P, Downing AK, Rao Z, Hewett DR, Sykes BC, Kielty CM. The calcium binding properties and molecular organization of epidermal growth factor-like domains in human fibrillin-1. *J Biol Chem* 1995;270(12):6751–6. <https://doi.org/10.1074/jbc.270.12.6751>
 - [55] Stenflo J, Stenberg Y, Muranyi A. Calcium-binding EGF-like modules in coagulation proteinases: function of the calcium ion in module interactions. *Biochim Biophys Acta* 2000;1477(1–2):51–63. [https://doi.org/10.1016/S0167-4838\(99\)00262-9](https://doi.org/10.1016/S0167-4838(99)00262-9)
 - [56] Fuentes-Prior P, Iwanaga Y, Huber R, Pagila R, Rumennik G, Seto M, et al. Structural basis for the anticoagulant activity of the thrombin-thrombomodulin complex. *Nature* 2000;404(6777):518–25. <https://doi.org/10.1038/35006683>
 - [57] Ranucci M, Sitzia C, Baryshnikova E, Di Dedda U, Cardani R, Martelli F, et al. Covid-19-associated coagulopathy: biomarkers of thrombin generation and fibrinolysis leading the outcome. *J Clin Med* 2020;9(11):3487. <https://doi.org/10.3390/jcm9113487>
 - [58] Goshua G, Pine AB, Meizlish ML, Chang C-H, Zhang H, Bahel P, et al. Endotheliopathy in COVID-19-associated coagulopathy: evidence from a single-centre, cross-sectional study. *Lancet Haematol* 2020;7(8):e575–82. [https://doi.org/10.1016/S2352-3026\(20\)30216-7](https://doi.org/10.1016/S2352-3026(20)30216-7)
 - [59] Bongiovanni D, Klug M, Lazareva O, Weidlich S, Biasi M, Ursu S, et al. SARS-CoV-2 infection is associated with a pro-thrombotic platelet phenotype. *Cell Death Dis* 2021;12(1). <https://doi.org/10.1038/s41419-020-03333-9>
 - [60] Freed DM, Bessman NJ, Kiyatkin A, Salazar-Cavazos E, Byrne PO, Moore JO, et al. EGF-like ligands differentially stabilize receptor dimers to specify signaling kinetics. *Cell* 2017;171(3):683–695.e18. <https://doi.org/10.1016/j.cell.2017.09.017>
 - [61] Garrett TPJ, McKern NM, Lou M, Elleman TC, Adams TE, Lovrecz GO, et al. Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor. *Cell* 2002;110:763–73. [https://doi.org/10.1016/S0092-8674\(02\)00940-6](https://doi.org/10.1016/S0092-8674(02)00940-6)
 - [62] Alcamí A, Smith GL. A soluble receptor for interleukin-1 β encoded by vaccinia virus: a novel mechanism of virus modulation of the host response to infection. *Cell* 1992;71(1):153–67. [https://doi.org/10.1016/0092-8674\(92\)90274-C](https://doi.org/10.1016/0092-8674(92)90274-C)
 - [63] Ahuja SK, Murphy PM. In: *Chemokines in Disease*. Totowa, NJ: Humana Press; 1999. p. 235–51. https://doi.org/10.1007/978-1-59259-706-2_15
 - [64] Chen I-Y, Chang SC, Wu H-Y, Yu T-C, Wei W-C, Lin S, et al. Upregulation of the chemokine (C-C Motif) ligand 2 via a severe acute respiratory syndrome coronavirus spike-ACE2 signaling pathway. *J Virol* 2010;84(15):7703–12. <https://doi.org/10.1128/JVI.02560-09>
 - [65] Zaid Y, Puhm F, Allaes I, Naya A, Oudghiri M, Khalki L, et al. Platelets Can associate with SARS-CoV-2 RNA and are hyperactivated in COVID-19. *Circ Res* 2020;127(11):1404–18. <https://doi.org/10.1161/CIRCRESAHA.120.317703>
 - [66] Jamilloux Y, Henry T, Belot A, Viel S, Fauter M, El T, et al. Should we suppress or stimulate immune responses for covid-19. *Autoimmun Rev* 2020;19.
 - [67] Buszko M, Park J-H, Verthelyi D, Sen R, Young HA, Rosenberg AS. The dynamic changes in cytokine responses in COVID-19: a snapshot of the current state of knowledge. *Nat Immunol* 2020;21(10):1146–51. <https://doi.org/10.1038/s41590-020-0779-1>
 - [68] Zizzo G, Cohen PL. Imperfect storm: is interleukin-33 the Achilles heel of COVID-19? *Lancet Rheumatol* 2020;2(12):e779–90. [https://doi.org/10.1016/S2665-9913\(20\)30340-4](https://doi.org/10.1016/S2665-9913(20)30340-4)
 - [69] Alsamman AM, Zayed H, Schindler M. The transcriptomic profiling of SARS-CoV-2 compared to SARS, MERS, EBOV, and H1N1. *PLoS ONE* 2020;15(12):e0243270.
 - [70] Beenken A, Eliseenkova AV, Ibrahim OA, Olsen SK, Mohammadi M. Plasticity in interactions of fibroblast growth factor 1 (FGF1) N terminus with FGF receptors underlies promiscuity of FGF1. *J Biol Chem* 2012;287:3067–78. <https://doi.org/10.1074/jbc.M111.275891>
 - [71] Xu K, Tzvetkova-Robev D, Xu Y, Goldgur Y, Chan Y-P, Himanen JP, et al. Insights into Eph receptor tyrosine kinase activation from crystal structures of the EphA4 ectodomain and its complex with ephrin-A5. *Proc Natl Acad Sci* 2013;110:14634 LP – 14639. <https://doi.org/10.1073/pnas.1311000110>
 - [72] Bowden TA, Aricescu AR, Nettleship JE, Siebold C, Rahman-Huq N, Owens RJ, et al. Structural plasticity of eph receptor A4 facilitates cross-class ephrin signaling. *Structure* 2009;17(10):1386–97.
 - [73] Wang J, Zheng X, Peng Q, Zhang X, Qin Z. Eph receptors: the bridge linking host and virus. *Cell Mol Life Sci* 2020;77(12):2355–65. <https://doi.org/10.1007/s00118-019-03409-6>
 - [74] Denorme F, Vanhoorelbeke K, De Meyer SF, von willebrand factor and platelet glycoprotein Ib: A thromboinflammatory axis in stroke. *Front Immunol* 2019;10:2884.
 - [75] Klok FA, Kruij MJHA, van der Meer NJM, Arbous MS, Gommers DAMPJ, Kant KM, et al. Incidence of thrombotic complications in critically ill ICU patients with COVID-19. *Thromb Res* 2020;191:145–7. <https://doi.org/10.1016/j.thromres.2020.04.013>
 - [76] Klammert U, Mueller TD, Hellmann TV, Wuerzler KK, Kotzsch A, Schliermann A, et al. GDF-5 can act as a context-dependent BMP-2 antagonist. *BMC Biol* 2015;13(1). <https://doi.org/10.1186/s12915-015-0183-8>
 - [77] Magro G. SARS-CoV-2 and COVID-19: Is interleukin-6 (IL-6) the ‘culprit lesion’ of ARDS onset? What is there besides tocilizumab? *SGP130Fc*. *Cytokine X* 2020;2(2):100029. <https://doi.org/10.1016/j.cytocx.2020.100029>
 - [78] Chen LYC, Hoiland RL, Stukas S, Wellington CL, Sekhon MS. Confronting the controversy: interleukin-6 and the COVID-19 cytokine storm syndrome. *Eur Respir J* 2020;56(4):2003006. <https://doi.org/10.1183/13993003.03006-2020>
 - [79] Zhang X, Wu K, Wang D, Yue X, Song D, Zhu Y, et al. Nucleocapsid protein of SARS-CoV activates interleukin-6 expression through cellular transcription factor NF-kappaB. *Virology* 2007;365:324–35. <https://doi.org/10.1016/j.virol.2007.04.009>
 - [80] Mendoza JL, Schneider WM, Hoffmann H-H, Vercauteren K, Jude KM, Xiong A, et al. The IFN- λ -IFN- λ R1-IL-10R β complex reveals structural features underlying type III IFN functional plasticity. *Immunity* 2017;46(3):379–92. <https://doi.org/10.1016/j.immuni.2017.02.017>
 - [81] Corbett AJ, Eckle SBG, Birkinshaw RW, Liu L, Patel O, Mahony J, et al. T-cell activation by transitory neo-antigens derived from distinct microbial pathways. *Nature* 2014;509(7500):361–5. <https://doi.org/10.1038/nature13160>
 - [82] Yu G, Boone T, Delaney J, Hawkins N, Kelley M, Ramakrishnan M, et al. APRIL and TALL-1 and receptors BCMA and TACI: system for regulating humoral immunity. *Nat Immunol* 2000;1:252–6. <https://doi.org/10.1038/79802>
 - [83] Tamada K, Shimozaki K, Chapoval AI, Zhai Y, Su J, Chen S-F, et al. LIGHT, a TNF-like molecule, costimulates T cell proliferation and is required for dendritic cell-mediated allogeneic T cell response. *J Immunol* 2000;164(8):4105–10. <https://doi.org/10.4049/jimmunol.164.8.4105>
 - [84] Smart OS, Womack TO, Flensburg C, Keller P, Paciorek W, Sharff A, et al. Exploiting structure similarity in refinement: automated NCS and target-structure restraints in (|it BUSTER). *Acta Crystallogr Sect D* 2012;68:368–80. <https://doi.org/10.1107/S0907444911056058>
 - [85] Nonaka M, Yoshizaki F. Evolution of the complement system. *Mol Immunol* 2004;40:897–902. <https://doi.org/10.1016/j.molimm.2003.10.009>
 - [86] Bernet J, Mullick J, Singh AK, Sahu A. Viral mimicry of the complement system. *J Biosci* 2003;28(3):249–64.
 - [87] Java A, Apicelli AJ, Kathryn Liszewski M, Coler-Reilly A, Atkinson JP, Kim AHJ, et al. The complement system in COVID-19: Friend and foe? *JCI Insight* 2020;5:1–12. <https://doi.org/10.1172/jci.insight.140711>
 - [88] Magro C, Mulvey JJ, Berlin D, Nuovo G, Salvatore S, Harp J, et al. Complement associated microvascular injury and thrombosis in the pathogenesis of severe

- COVID-19 infection: a report of five cases. *Transl Res* 2020;220:1–13. <https://doi.org/10.1016/j.trsl.2020.04.007>.
- [89] Ramlall V, Thangaraj PM, Meydan C, Foox J, Butler D, May B, et al. Identification of Immune complement function as a determinant of adverse SARS-CoV-2 infection outcome. *MedRxiv* 2020. <https://doi.org/10.1101/2020.05.05.20092452>.
- [90] Yu J, Yuan X, Chen H, Chaturvedi S, Braunstein EM, Brodsky RA. Direct activation of the alternative complement pathway by SARS-CoV-2 spike proteins is blocked by factor D inhibition. *Blood* 2020;136:2080–9. <https://doi.org/10.1182/blood.202008248>.
- [91] Xue X, Wu J, Ricklin D, Forneris F, Di Crescenzo P, Schmidt CQ, et al. Regulator-dependent mechanisms of C3b processing by factor I allow differentiation of immune responses. *Nat Struct Mol Biol* 2017;24(8):643–51. <https://doi.org/10.1038/nsmb.3427>.
- [92] van den Elsen JMH, Isenman DE. A crystal structure of the complex between human complement receptor 2 and its ligand C3d. *Science* (80-) 2011;332(6029):608–11. <https://doi.org/10.1126/science.1201954>.
- [93] Campbell ID, Spitzfaden C. Building proteins with fibronectin type III modules. *Structure* 1994;2(5):333–7. [https://doi.org/10.1016/S0969-2126\(00\)00034-4](https://doi.org/10.1016/S0969-2126(00)00034-4).
- [94] Leahy DJ, Aukhil I, Erickson HP. Crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. *Cell* 1996;84:155–64. [https://doi.org/10.1016/S0092-8674\(00\)81002-8](https://doi.org/10.1016/S0092-8674(00)81002-8).
- [95] Mendiratta SS, Sekulic N, Hernandez-Guzman FG, Close BE, Lavie A, Colley KJ. A novel α -helix in the first fibronectin type III repeat of the neural cell adhesion molecule is critical for N-glycan polysialylation. *J Biol Chem* 2006;281:36052–9. <https://doi.org/10.1074/jbc.M608073200>.
- [96] Luca VC, Kim BC, Ge C, Kakuda S, Wu Di, Rooin-Peikar M, et al. Notch-Jagged complex structure implicates a catch bond in tuning ligand sensitivity. *Science* 2017;355(6331):1320–4. <https://doi.org/10.1126/science.aaf9739>.
- [97] Renko M, Sabotić J, Mihelić M, Brzin J, Kos J, Turk D. Versatile loops in mycocyprins inhibit three protease families. *J Biol Chem* 2010;285:308–16. <https://doi.org/10.1074/jbc.M109.043331>.
- [98] Pišlar A, Mitrović A, Sabotić J, Pečar Fonović U, Perišić Nanut M, Jakoš T, et al. The role of cysteine peptidases in coronavirus cell entry and replication: the therapeutic potential of cathepsin inhibitors. *PLOS Pathog* 2020;16(11):e1009013.
- [99] Cheng Y, Zak O, Aisen P, Harrison SC, Walz T. Structure of the human transferrin receptor-transferrin complex. *Cell* 2004;116(4):565–76. [https://doi.org/10.1016/S0092-8674\(04\)00130-8](https://doi.org/10.1016/S0092-8674(04)00130-8).
- [100] Watson AA, Lebedev AA, Hall BA, Fenton-May AE, Vagin AA, Dejnirattisai W, et al. Structural flexibility of the macrophage dengue virus receptor CLEC5A: implications for ligand binding and signaling. *J Biol Chem* 2011;286(27):24208–18.
- [101] Zhang H, Casasnovas JM, Jin M, Liu J-H, Gahmberg CG, Springer TA, et al. An unusual allosteric mobility of the C-terminal helix of a high-affinity α integrin i domain variant bound to ICAM-5. *Mol Cell* 2008;31(3):432–7. <https://doi.org/10.1016/j.molcel.2008.06.022>.
- [102] Changeux J-P, Amoura Z, Rey FA, Miyara M. A nicotinic hypothesis for Covid-19 with preventive and therapeutic implications. *C R Biol* 2020;343:33–9. <https://doi.org/10.5802/crbil.8>.
- [103] Pizarro JC, Chitarra V, Verger D, Holm I, Pêtres S, Darteville S, et al. Crystal structure of a fab complex formed with PfMSP1-19, the C-terminal fragment of merozoite surface protein 1 from *Plasmodium falciparum*: a malaria vaccine candidate. *J Mol Biol* 2003;328(5):1091–103.
- [104] Igonet S, Vulliez-Le Normand B, Faure G, Riottot M-M, Kocken CHM, Thomas AW, et al. Cross-reactivity studies of an anti-plasmodium vivax apical membrane antigen 1 monoclonal antibody: binding and structural characterisation. *J Mol Biol* 2007;366(5):1523–37.
- [105] Craig L, Sanschagrin PC, Rozek A, Lackie S, Kuhn IA, Scott JK. The role of structure in antibody cross-reactivity between peptides and folded proteins. *J Mol Biol* 1998;281:183–201. <https://doi.org/10.1006/jmbi.1998.1907>.
- [106] Panda AK, Tripathy R, Das BK. *Plasmodium falciparum* infection may protect a population from severe acute respiratory syndrome coronavirus 2 infection. *J Infect Dis* 2020;222:1570–1. <https://doi.org/10.1093/infdis/jiaa455>.
- [107] Iesa MAM, Osman MEM, Hassan MA, Dirar AIA, Abuzeid N, Mancuso JJ, et al. SARS-CoV-2 and plasmodium falciparum common immunodominant regions may explain low COVID-19 incidence in the malaria-endemic belt. *New Microbes* 2020;38:100817. <https://doi.org/10.1016/j.nmni.2020.100817>.
- [108] Kalungi A, Kinyanda E, Akena DH, Kaleebu P, Bisangwa IM. Less severe cases of COVID-19 in sub-saharan africa: could co-infection or a recent history of plasmodium falciparum infection be protective? *Front Immunol* 2021;12:6360. <https://doi.org/10.3389/fimmu.2021.565625>.
- [109] Raham TF. Influence of malaria endemicity and tuberculosis prevalence on COVID-19 mortality. *Public Health* 2021;194:33–5. <https://doi.org/10.1016/j.puhe.2021.02.018>.
- [110] Jankowiak Ł, Rozsa L, Tryjanowski P, Møller AP. A negative covariation between toxoplasmosis and CoVID-19 with alternative interpretations. *Sci Rep* 2020;10:12512. <https://doi.org/10.1038/s41598-020-69351-x>.
- [111] Ploss A, Evans MJ. Hepatitis C virus host cell entry. *Curr Opin Virol* 2012;2(1):14–9. <https://doi.org/10.1016/j.coviro.2011.12.007>.
- [112] Heile JM, Fong Y-L, Rosa D, Berger K, Saletti G, Campagnoli S, et al. Evaluation of Hepatitis C Virus Glycoprotein E2 for Vaccine Design: an Endoplasmic Reticulum-Retained Recombinant Protein Is Superior to Secreted Recombinant Protein and DNA-Based Vaccine Candidates. *J Virol* 2000;74:6885 LP – 6892. <https://doi.org/10.1128/JVI.74.15.6885-6892.2000>.
- [113] Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City Area. *JAMA* 2020;323(20):2052. <https://doi.org/10.1001/jama.2020.6775>.
- [114] Reddy KR. SARS-CoV-2 and the liver: considerations in hepatitis b and hepatitis C infections. *Clin Liver Dis* 2020;15:191–4. <https://doi.org/10.1002/clid.970>.
- [115] Mirzaie H, Vahidi M, Shokoohi M, Darvishian M, Sharifi H, Sharafi H, et al. COVID-19 among patients with hepatitis B or hepatitis C: A systematic review. *MedRxiv* 2020:2020.10.22.20216317. <https://doi.org/10.1101/2020.10.22.20216317>.
- [116] Brockbank SM V, Soden J, Faba-Rodriguez R, Ribeiro LR, Geh C, Thomas H, et al. SARS-CoV-2 comprehensive receptor profiling: mechanistic insight to drive new therapeutic strategies. *BioRxiv* 2021:2021.03.11.434937. <https://doi.org/10.1101/2021.03.11.434937>.
- [117] Abraham J, Corbett KD, Farzan M, Choe H, Harrison SC. Structural basis for receptor recognition by New World hemorrhagic fever arenaviruses. *Nat Struct Mol Biol* 2010;17(4):438–44. <https://doi.org/10.1038/nsmb.1772>.
- [118] Moonens K, Gideonsson P, Subedi S, Bugaytsova J, Romão E, Mendez M, et al. Structural insights into polymorphic ABO glycan binding by helicobacter pylori. *Cell Host Microbe* 2016;19(1):55–66.
- [119] Saez NJ, Mobli M, Bieri M, Chassagnon IR, Malde AK, Gamsjaeger R, et al. A dynamic pharmacophore drives the interaction between Psalmotoxin-1 and the putative drug target acid-sensing ion channel 1a. *Mol Pharmacol* 2011;80(5):796–808. <https://doi.org/10.1124/mol.111.072207>.
- [120] Volpone L, Lamthanh H, Barbier J, Gilles N, Molgó J, Ménez A, et al. NMR solution structures of δ -conotoxin EVIA from *Conus ermineus* that selectively acts on vertebrate neuronal Na⁺ channels. *J Biol Chem* 2004;279(2):21356–66.
- [121] Dellisanti CD, Yao Y, Stroud JC, Wang Z-Z, Chen L. Crystal structure of the extracellular domain of nAChR α 1 bound to α -bungarotoxin at 1.94 Å resolution. *Nat Neurosci* 2007;10:953–62. <https://doi.org/10.1038/nn1942>.
- [122] Lee S-C, Lin C-C, Wang C-H, Wu P-L, Huang H-W, Chang C-I, et al. Endocytotic routes of cobra cardiotoxins depend on spatial distribution of positively charged and hydrophobic domains to target distinct types of sulfated glycoconjugates on cell surface. *J Biol Chem* 2014;289(29):20170–81. <https://doi.org/10.1074/jbc.M114.557157>.
- [123] Corfield PW, Lee TJ, Low BW. The crystal structure of erabutoxin a at 2.0-Å resolution. *J Biol Chem* 1989;264:9239–42. <https://doi.org/10.2210/pdb5ebx/pdb>.
- [124] McFarland AJ, Yousuf MS, Shiers S, Price TJ. Neurobiology of SARS-CoV-2 interactions with the peripheral nervous system: implications for COVID-19 and pain. *PAIN Rep* 2021;6(1):e885. <https://doi.org/10.1097/PR9.0000000000000885>.
- [125] Moutal A, Martin LF, Boiron L, Gomez K, Ran D, Zhou Y, et al. SARS-CoV-2 Spike protein co-opts VEGF-A/Neuropilin-1 receptor signaling to induce analgesia. *Pain* 2021;162(1):243–52. <https://doi.org/10.1097/j.pain.0000000000002097>.
- [126] Iba T, Levy JH, Levi M, Connors JM, Thachil J. Coagulopathy of coronavirus disease 2019. *Crit Care Med* 2020;48:1358–64. <https://doi.org/10.1097/CCM.0000000000004458>.
- [127] Parvizpour S, Pourseif MM, Razmara J, Rafi MA, Omid Y. Epitope-based vaccine design: a comprehensive overview of bioinformatics approaches. *Drug Discov Today* 2020;25:1034–42. <https://doi.org/10.1016/j.drudis.2020.03.006>.
- [128] Mastellos D, Morikis D, Isaacs SN, Holland MC, Strey CW, Lambris JD. Complement: structure, functions, evolution, and viral molecular mimicry. *Immunol Res* 2003;27(2–3):367–86. <https://doi.org/10.1385/IR:27-2-3:367>.
- [129] Rowan EG. What does beta-bungarotoxin do at the neuromuscular junction? *Toxicol* 2001;39:107–18. [https://doi.org/10.1016/S0041-0101\(00\)00159-8](https://doi.org/10.1016/S0041-0101(00)00159-8).
- [130] Dongol Y, Caldas Cardoso F, Lewis RJ. Spider knottin pharmacology at voltage-gated sodium channels and their potential to modulate pain pathways. *Toxins* (Basel) 2019;11(11):626. <https://doi.org/10.3390/toxins11110626>.
- [131] Wu P-L, Lee S-C, Chuang C-C, Mori S, Akakura N, Wu W, et al. Non-cytotoxic cobra cardiotoxin A5 binds to α (v)beta3 integrin and inhibits bone resorption. Identification of cardiotoxins as non-RGD integrin-binding proteins of the Ly-6 family. *J Biol Chem* 2006;281:7937–45. <https://doi.org/10.1074/jbc.M513035200>.
- [132] Cao L, Goresnik I, Coventry B, Case JB, Miller L, Kozodoy L, et al. De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science* (80-) 2020;370:426 LP – 431. <https://doi.org/10.1126/science.abd9909>.
- [133] Brister JR, Ako-Adjei D, Bao Y, Blinkova O. NCBI viral genomes resource. *Nucleic Acids Res* 2015;43:D571–7. <https://doi.org/10.1093/nar/gku1207>.
- [134] Torres PHM, Rossi AD, Blundell TL. ProtCHOIR: a tool for proteome-scale generation of homo-oligomers. *Brief Bioinform* 2021. <https://doi.org/10.1093/bib/bbab182>.
- [135] Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 1993;234(3):779–815.
- [136] Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 2007;372(3):774–97. <https://doi.org/10.1016/j.jmb.2007.05.022>.

- [137] Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–402.
- [138] Krissinel E. Enhanced fold recognition using efficient short fragment clustering. *J Mol Biochem* 2012;1:76–85.
- [139] Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr Sect D Biol Crystallogr* 2010;66(1):12–21. <https://doi.org/10.1107/S0907444909042073>.
- [140] Šali A. MODELLER: A Program for Protein Structure Modeling 2019:779–815.
- [141] Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* 2019;47:W636–41. <https://doi.org/10.1093/nar/gkz268>.
- [142] Alsulami AF, Thomas SE, Jamasb AR, Beaudoin CA, Moghul I, Bannerman B, et al. SARS-CoV-2 3D database: understanding the coronavirus proteome and evaluating possible drug targets. *Brief Bioinform* 2021;22:769–80. <https://doi.org/10.1093/bib/bbaa404>.
- [143] Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager K-B G, et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J* 2013;32(10):1478–88. <https://doi.org/10.1038/emboj.2013.79>.
- [144] Gupta R, Jung E, Brunak S. Prediction of N-glycosylation sites in human proteins. In preparation. 2004.
- [145] Kuriata A, Gierut AM, Oleniecki T, Ciemny MP, Kolinski A, Kurcinski M, et al. CABS-flex 2.0: a web server for fast simulations of flexibility of protein structures. *Nucleic Acids Res* 2018;46:W338–43. <https://doi.org/10.1093/nar/gky356>.
- [146] Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A. Coarse-grained protein models and their applications. *Chem Rev* 2016;116(14):7898–936. <https://doi.org/10.1021/acs.chemrev.6b00163>.
- [147] Holm L, Laakso LM. Dali server update. *Nucleic Acids Res* 2016;44(W1):W351–5. <https://doi.org/10.1093/nar/gkw357>.
- [148] Li Z, Jaroszewski L, Iyer M, Sedova M, Godzik A. FATCAT 2.0: towards a better understanding of the structural diversity of proteins. *Nucleic Acids Res* 2020;48:W60–4. <https://doi.org/10.1093/nar/gkaa443>.
- [149] Lo W-C, Lee C-Y, Lee C-C, Lyu P-C. iSARST: an integrated SARST web server for rapid protein structural similarity searches. *Nucleic Acids Res* 2009;37:W545–51. <https://doi.org/10.1093/nar/gkp291>.
- [150] Deng L, Zhong G, Liu C, Luo J, Liu H. MADOKA: An ultra-fast approach for large-scale protein structure similarity searching. *BMC Bioinf* 2019;20:1–10. <https://doi.org/10.1186/s12859-019-3235-1>.
- [151] Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr Sect D Biol Crystallogr* 2004;60(12):2256–68. <https://doi.org/10.1107/S0907444904026460>.
- [152] Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 2005;33:2302–9. <https://doi.org/10.1093/nar/gki524>.
- [153] Ayoub R, Lee Y, Zhang Y. Rupee: A fast and accurate purely geometric protein structure search. *PLoS ONE* 2019;14(3):e0213712. <https://doi.org/10.1371/journal.pone.0213712>.
- [154] Dong R, Pan S, Peng Z, Zhang Y, Yang J. MTM-align: a server for fast protein structure database search and multiple protein structure alignment. *Nucleic Acids Res* 2018;46:W380–6. <https://doi.org/10.1093/nar/gky430>.
- [155] Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983;22(12):2577–637. <https://doi.org/10.1002/bip.360221211>.
- [156] Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. *Proteins* 2004;57(4):702–10. <https://doi.org/10.1002/prot.20264>.
- [157] Hutchinson EG, Thornton JM. PROMOTIF—a program to identify and analyze structural motifs in proteins. *Protein Sci* 1996;5(2):212–20. <https://doi.org/10.1002/pro.5560050204>.
- [158] Krissinel E, Henrick K. Multiple Alignment of Protein Structures in Three Dimensions BT - Computational Life Sciences. In: R. Berthold M, Glen RC, Diederichs K, Kohlbacher O, Fischer I, editors., Berlin, Heidelberg: Springer Berlin Heidelberg; 2005, p. 67–78.
- [159] Sillitoe I, Dawson N, Lewis TE, Das S, Lees JG, Ashford P, et al. CATH: Expanding the horizons of structure-based functional annotations for genome sequences. *Nucleic Acids Res* 2019;47:D280–4. <https://doi.org/10.1093/nar/gky1097>.
- [160] Andreeva A, Kulesha E, Gough J, Murzin AG. The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures. *Nucleic Acids Res* 2020;48:D376–82. <https://doi.org/10.1093/nar/gkz1064>.
- [161] Cheng H, Schaeffer RD, Liao Y, Kinch LN, Pei J, Shi S, et al. ECOD: an evolutionary classification of protein domains. *PLoS Comput Biol* 2014;10(12):e1003926. <https://doi.org/10.1371/journal.pcbi.1003926>.
- [162] Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Di Costanzo L, et al. RCSB Protein Data Bank: Biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res* 2019;47:D464–74. <https://doi.org/10.1093/nar/gky1004>.
- [163] Guven-Maiorov E, Hakouz A, Valjevac S, Keskin O, Tsai C-J, Gursoy A, et al. HMI-PRED: a web server for structural prediction of host-microbe interactions based on interface mimicry. *J Mol Biol* 2020;432(11):3395–403. <https://doi.org/10.1016/j.jmb.2020.01.025>.
- [164] Hubbard S, Thornton J. NACCESS, Department of Biochemistry Molecular Biology, University College London 1993.
- [165] Wang C, Bradley P, Baker D. Protein-protein docking with backbone flexibility. *J Mol Biol* 2007;373(2):503–19. <https://doi.org/10.1016/j.jmb.2007.07.050>.
- [166] Bateman A. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2019;47:D506–15. <https://doi.org/10.1093/nar/gky1049>.
- [167] Kozakov D, Hallc DR, Xiab B, Porterb KA, Padhornya D, Yuehb C, et al. The ClusPro web server for protein-protein docking. *Nat Protoc* 2017;12:255–78. <https://doi.org/10.1038/nprot.2016.169>. The.
- [168] Vajda S, Yueh C, Beglov D, Bohnuud T, Mottarella SE, Xia B, et al. New additions to the ClusPro server motivated by CAPRI. *Proteins Struct Funct Bioinforma* 2017;85:435–44. <https://doi.org/10.1002/prot.25219>.
- [169] Desta IT, Porter KA, Xia B, Kozakov D, Vajda S. Performance and Its Limits in Rigid Body Protein-Protein Docking. *Structure* 2020;28(9):1071–1081.e3. <https://doi.org/10.1016/j.str.2020.06.006>.
- [170] Kozakov D, Beglov D, Bohnuud T, Mottarella SE, Xia B, Hall DR, et al. How good is automated protein docking? *Proteins* 2013;81:2159–66. <https://doi.org/10.1002/prot.24403>. How.
- [171] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 2019;47:D607–13. <https://doi.org/10.1093/nar/gky1131>.
- [172] Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucleic Acids Res* 2005;33:W382–8. <https://doi.org/10.1093/nar/gki387>.
- [173] Stout T. Show charged amino acids on PyMOL 2004.
- [174] Rodrigues J, Teixeira JMC, Trellet M, Bonvin A. pdb-tools: a swiss army knife for molecular structures. *F1000Research* 2018;7. <https://doi.org/10.12688/f1000research.17456.1>.