Genes & Immunity (2020) 21:409–419
https://doi.org/10.1038/s41435-020-00120-6

ARTICLE

Evolutionary and structural analysis of SARS-CoV-2 specific evasion of host immunity

Irfan Hussain[1] · Nashaiman Pervaiz[1] · Abbas Khan[2] · Shoaib Saleem[1] · Huma Shireen[1] · Dong-Qing Wei[2,3,4] · Viviane Labrie[5,6] · Yiming Bao[7,8] · Amir Ali Abbasi[1]

Received: 6 June 2020 / Revised: 5 November 2020 / Accepted: 6 November 2020 / Published online: 3 December 2020

Abstract

The outbreak of coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is spreading fast worldwide. There is a pressing need to understand how the virus counteracts host innate immune responses. Deleterious clinical manifestations of coronaviruses have been associated with virus-induced direct dysregulation of innate immune responses occurring via viral macrodomains located within nonstructural protein-3 (Nsp3). However, no substantial information is available concerning the relationship of macrodomains to the unusually high pathogenicity of SARS-CoV-2. Here, we show that structural evolution of macrodomains may impart a critical role to the unique pathogenicity of SARS-CoV-2. Using sequence, structural, and phylogenetic analysis, we identify a specific set of historical substitutions that recapitulate the evolution of the macrodomains that counteract host immune response. These evolutionary substitutions may alter and reposition the secondary structural elements to create new intra-protein contacts and, thereby, may enhance the ability of SARS-CoV-2 to inhibit host immunity. Further, we find that the unusual virulence of this virus is potentially the consequence of Darwinian selection-driven epistasis in protein evolution. Our findings warrant further characterization of macrodomain-specific evolutionary substitutions in in vitro and in vivo models to determine their inhibitory effects on the host immune system.

Introduction

Since the first reports of patients with atypical pneumonia or coronavirus disease 2019 (COVID-19) in Wuhan, China in late December 2019, the outbreak has now become a

These authors contributed equally: Irfan Hussain, Nashaiman Pervaiz, Abbas Khan

Supplementary information The online version of this article (https://doi.org/10.1038/s41435-020-00120-6) contains supplementary material, which is available to authorized users.

✉ Dong-Qing Wei
dqwei@sjtu.edu.cn

✉ Yiming Bao
baoym@big.ac.cn

✉ Amir Ali Abbasi
abbasiam@qau.edu.pk

[1] National Center for Bioinformatics, Program of Comparative and Evolutionary Genomics, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan

[2] State Key Lab of Microbial Metabolism, Department of Bioinformatics and Biological Statistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 200240 Shanghai, PR China

[3] State Key Laboratory of Microbial Metabolism, Shanghai-Islamabad-Belgrade Joint Innovation Center on Antibacterial Resistances, Joint Laboratory of International Cooperation in Metabolic and Developmental Sciences, Ministry of Education and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 200030 Shanghai, PR China

[4] Peng Cheng Laboratory, Vanke Cloud City, Phase I Building 8, Xili Street, Nashan District, Guangdong 518055 Shenzhen, PR China

[5] Center for Neurodegenerative Science, Van Andel Research Institute, Grand Rapids, MI 49503, USA

[6] Division of Psychiatry and Behavioral Medicine, College of Human Medicine, Michigan State University, Grand Rapids, MI 49503, USA

[7] National Genomics Data Center & CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, China National Center for Bioinformation, 100101 Beijing, PR China

[8] University of Chinese Academy of Sciences, 100101 Beijing, PR China

pandemic with global socioeconomic impact. Presently, there is no vaccine or specific antiviral treatment for COVID-19. On 7 January 2020, through genome sequencing technology, a novel coronavirus (nCoV) was identified as the causative pathogen, named as 2019-nCoV (also referred as SARS-CoV-2) [1].

CoVs are commonly associated with respiratory and gastrointestinal tract infections and constitute a phylogenetically diverse viral group, comprising of four genera: *alphacoronavirus* (α-CoV), *betacoronavirus* (β-CoV), *gammacoronavirus* (γ-CoV), and *deltacoronavirus* (δ-CoVs) [2]. They are complex pathogens that are known to infect multiple host species, including humans [3, 4]. Before the emergence of COVID-19, six CoVs were known to infect humans. For instance, HCoV-NL63 (β-CoV), HCoV-229E (β-CoV), HCoV-OC43 (α-CoV), and HKU1 (α-CoV) can cause mild upper respiratory infections, whereas SARS-CoV (β-CoV) and MERS-CoV (β-CoV) can infect the lower respiratory tract of humans and cause severe respiratory syndrome [5]. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) belonging to the group of β-CoVs, is the seventh CoV to infect humans and the third β-CoV to infect the lower respiratory tract. The mortality rate of SARS (9.6%) and MERS (34%) is reportedly higher than that of COVID-19 (3–6%) [6]. However, the spread of SARS-CoV-2 infection is remarkably wide and rapid [7].

Comparative genomics has revealed that SARS-CoV-2 possesses a genome architecture typical of CoVs, comprising of a ~29.8 kilobase (kb) single-stranded positive-sense RNA (+ssRNA) that contains 14 ORFs encoding for 27 proteins. The 5′-terminus of the genome encodes for two long polyproteins, pp1ab (7096 amino acids) and pp1a (4405 amino acids) [1]. These polyproteins are processed by virally encoded proteases to produce 10 nonstructural proteins (nsp1–nsp10), and, in addition, pp1ab uniquely produces nsp13 to nsp16 and pp1a uniquely produces nsp11. The 3′-terminus of the SARS-CoV-2 genome encodes for four main structural proteins: spike (S), membrane (M), envelope (E), nucleocapsid (N), and eight accessory proteins [8].

The availability of SARS-CoV-2 genome sequence data has initiated efforts to design diagnostic tests and potent therapies [9]. Furthermore, there is an urgent need to explore the evolutionary origin and phylogenetics of SARS-CoV-2 with the potential implication that it will further our understanding of disease pathogenesis and spread [1, 10, 11]. Traditionally, inference of evolutionary relationships among CoVs has relied heavily on comparisons of whole genome sequence data or of their critical structural genes, such as the *S* gene, which encodes the spike protein [1]. Nucleotide datasets comprised of multiple distinct genes exhibiting heterogeneity in their mode and rate of sequence evolution can adversely affect phylogenetic reconstruction [12]. Evading this problem is particularly challenging in CoVs. Comparative analysis has revealed a heterogeneous rate of evolution of portions of the genome of CoVs, with ~60% shared identity within nsp coding regions and only ~40% shared identity in the remaining one-third of the genome coding for structural proteins. This heterogeneity in substitution rates in coding regions of CoVs can potentially complicate attempts to reconstruct the evolutionary history of CoVs [13]. Specifically, the rapid rate of evolution of structural genes (such as the *S* gene) may cause a gene-based analysis to blur the history of the taxa [14]. In contrast, pp1ab is a large, slow-evolving domain in SARS-COV-2 capable of circumventing the potential pitfalls of using genomic-based approaches to reconstruct the history of SARS-CoV-2 [5]. Here, based on completely sequenced genomes of SARS-CoV-2, covering at least 39 distinct global territories, we conducted an in-depth comparative analysis of the 7096-aa replicase polyprotein pp1ab, comparing it to the corresponding homologous polyproteins of 83 related CoVs [15]. To best of our knowledge, this is the first attempt to use the full-length pp1ab polyprotein to predict SARS-CoV-2 relatedness to other members of the *Coronavirinae* subfamily. We next focused on macrodomains encoded within Nsp3 of pp1ab to investigate their structural evolution in light of previous associations between macrodomains and virulence, and the potential of these associations in designing a novel therapeutic strategy for the treatment of SARS-CoV-2 -induced severe infections.

Materials and methods

Sequence collection

The amino acid and coding sequences of ORF1ab gene from four genera of the subfamily Coronavirinae; Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus were retrieved from GenBank (http://www.ncbi.nlm.nih.gov) [16], 2019 Novel Coronavirus Resource of CNCB/NGDC [15, 17] and the GISAID databank (https://www.gisaid.org/) [18]. In total 247 sequences from four genera of subfamily Coronavirinae were used in this study (Supplementary Tables S1, S2; Supplementary Alignment File).

Phylogenetic and sequence analysis

The sequence alignment of 121 Coronavirinae polyprotein pp1ab sequences was performed using CLUSTALW (default parameters) [19]. The phylogenetic tree of the subfamily Coronavirinae, including SARS-CoV-2 was reconstructed in MEGA 5.05 by the maximum likelihood

(ML) method with the Whelan and Goldman (WAG) amino-acid substitution model [20, 21] The phylogenetic tree with the upmost log likelihood scores was selected. A neighbor-joining (NJ) tree was also reconstructed using uncorrected proportion (p) distance and the Jones, Taylor, and Thornton method (as amino-acid substitution model) to calculate evolutionary distances between coronaviruses [22, 23]. Topological reliability of the NJ and ML tree was tested by bootstrap analysis (1000 pseudoreplicates) [23].

All proteins encoded by the ORF1ab gene were aligned for SARS-CoV-2 (YP_009725299.1), bat-RaTG13 (QHR63299.1), bat-ZC45 (AVP78030.1), bat-ZXC21 (AVP78041.1), and SARS-CoVBJ01 (AAP30028.1) using MAFFT and Clustal Omega [24, 25]. Amino acids substitutions unique to SARS-CoV-2 were identified by manual inspection of the alignments (Supplementary Table S3). A sequence similarity plot was produced (sliding window with 5 amino acid-step) using the Plotcon software available in the EMBOSS software suite [26].

Thermodynamic state function, the ΔΔG (Gibbs free energy) of a substitution from ancestral protein to its altered version was predicted by employing the conformationally constrained environment-specific substitution tables [27]. The putative physicochemical impact of each substitution on protein structure and function was estimated using the BLOSUM-62 substitution matrix [28].

Ancestral sequence reconstruction

The ML method implemented in MEGA was used to reconstruct ancestral sequences of SARS-CoV-2 and bat-RaTG13/ZC45/ZXC21 based on amino-acid substitutions identified in the WAG model [20, 21, 29]. Separately, we also inferred the ancestral sequence using the PRANK program that accepted insertions and deletions as distinct evolutionary events [30]. The consensus of PRANK and MEGA ancestral sequence of SARS-CoV-2 and bat-RaTG13/ZC45/ZXC was used in this current study. Sequence alignments for ancestral reconstruction were performed using MUSCLE and MAFFT (default parameters) [24, 31]

Structural analysis

After the divergence from Bat-CoV-RaTG13, the SARS-CoV-2 nonstructural protein-3 (Nsp3) accumulated a greater number of substitutions compared to any other nonstructural protein encoded by the SARS-CoV-2 ORF1ab gene. To determine the functional effects of these Nsp3 substitutions in SARS-CoV-2, we performed a 3D structural analysis that examined all relevant protein structures using a homology modeling approach in Modeler [32, 33]. For the structural analysis, we based template selection on high sequence homology and amino-acid length/coverage and obtained the following templates from RCSB databank [34]: 2WCT, 2JZF, 2KQV, 2ACF, 6MEA, and 5DUS. Protein structures were predicted using the Discrete Optimized Protein Energy score, followed by implementation of the energy minimization protocol in PyMOL to maximally improve the quality of the modeled structures [35]. The quality of the predicted protein structures were further validated by Rampage Ramachandran plot analysis [36]. Superimposition of the modeled protein structures was performed with PyMOL, and root mean square deviation values were assessed [35, 37]. Furthermore, sequence-based secondary structure elements were determined using the PSIPRED server [38].

For comparative binding analysis of Mac-1 to ADPr, crystallographic structures of MERS-CoV (5DUS), SARS-CoV (2FAV), and SAR-CoV-2 (6W02) Mac-1 (within Nsp3) were obtained from the RCSB databank [34]. AutoDock [39] was used to perform an induce-fit docking (IFD) protocol with 30 conformers, while keeping the rest of the parameters default. IFD modeling offers mutual conformational adaptations of a protein receptor to a ligand, which enables for better accuracy than docking to a rigid target [40]. The best docking complexes were selected based on docking score. ADPr (ADP-ribose) and Mac-1 superimposition between CoVs were visualized in PyMOL. Interactions of the key residues of Mac-1 with ADPr protein was obtained and visualized in PyMOL. To confirm the differences in binding affinities of Mac-1 (for ADPr) of MERS-CoV, SARS-CoV, and SAR-CoV-2, DoGsitescorer (https://proteins.plus/) was used to calculate volume, surface area and druggability scores of binding cavities [41].

Results and discussion

ML and NJ trees exhibited similar topologies, where the cluster of δ-CoVs and γ-CoVs was the first to diverge, followed by α-CoVs, β-OC43-CoVs, β-MERS-CoVs, and β-SARS-CoVs, respectively (Fig. 1; Supplementary Figs. S1, S2). Tree topology confirms the direct grouping of SARS-CoV-2 with batCoV-RaTG13, which diverged from a cluster of batSLCoVZC45 and batSLCoVZXC21.This pattern places the SARS-CoV-2 and the three batSL-CoVs in a distinct phylogenetic group compare to SARS-CoVs and other SARS-like coronaviruses (Fig. 1).

Phylogenetic separation of SARS-CoV-2 from SARS-CoVs suggests that after their origin from the Hp-βCoV/BM48-31-like common ancestor, these two distinct lineages of CoVs were subjected to different genetic selection pressures. This may have led to notable differences in their infectivity, transmissibility, pathogenesis, and host tropism. This speculation prompted us to elucidate the putative

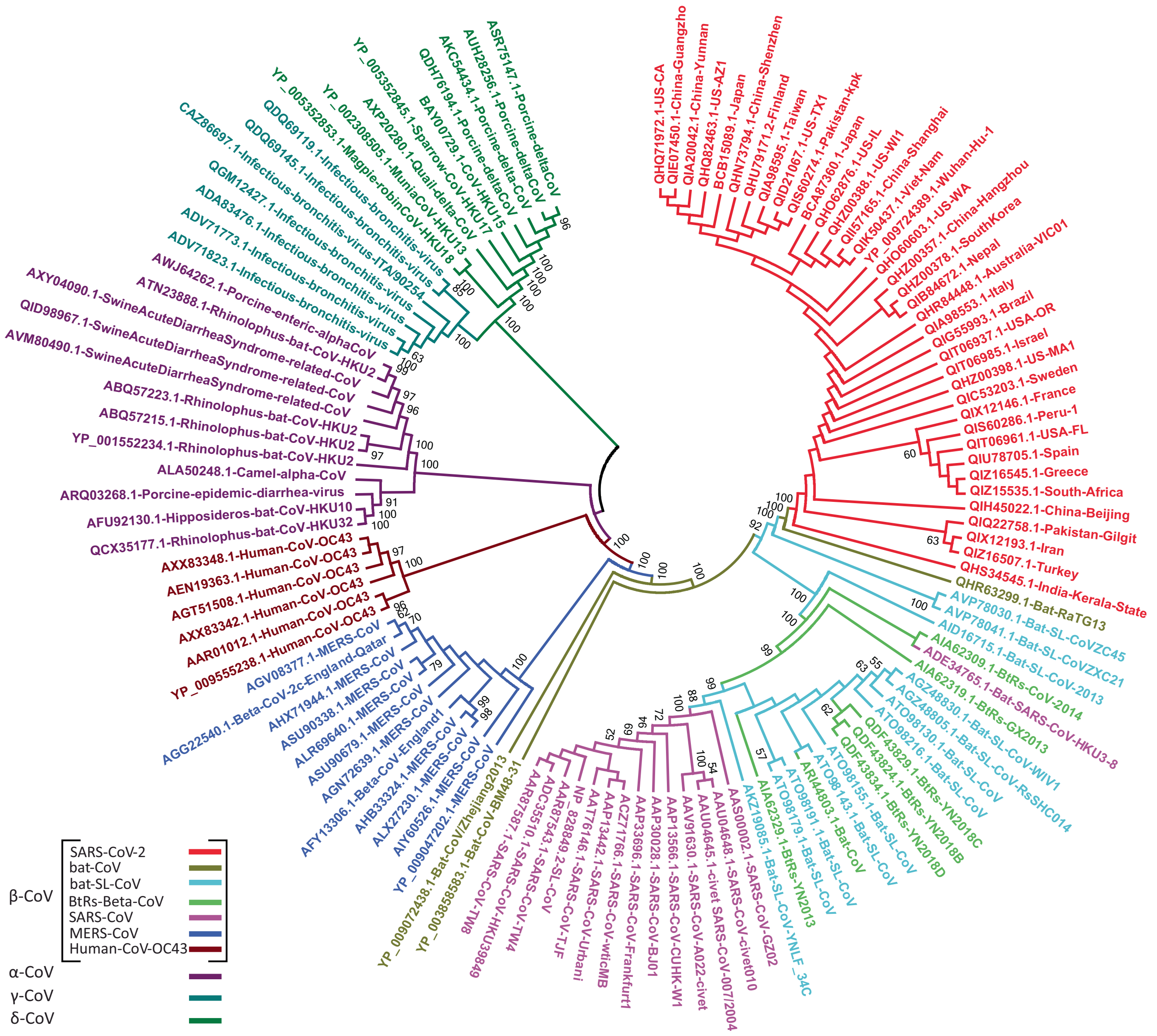
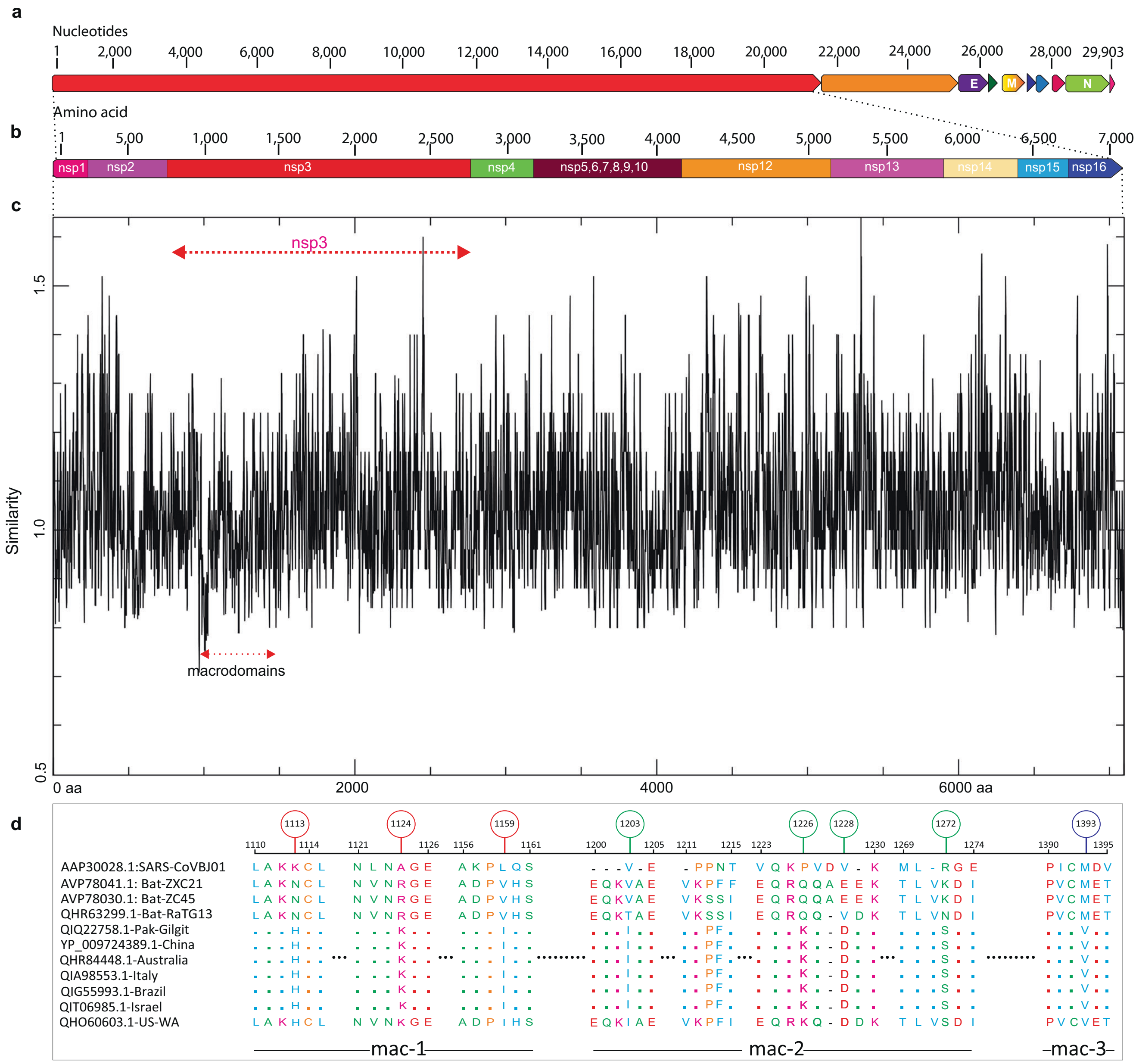

ASR75147.1-Porcine-deltaCoV
AUH28256.1-Porcine-deltaCoV
AKC54434.1-Porcine-deltaCoV
QDH76194.1-Porcine-deltaCoV
BAY00729.1-CoV-HKU15
YP_005352845.1-Sparrow-CoV-HKU17
AXP20280.1-Quail-delta-CoV
YP_002308505.1-MuniaCoV-HKU13
YP_005352853.1-Magpie-robinCoV-HKU18
QDQ69119.1-Infectious-bronchitis-virus
QDQ69145.1-Infectious-bronchitis-virus
CAZ86697.1-Infectious-bronchitis-virus-ITA/90254
QGM12427.1-Infectious-bronchitis-virus
ADA83476.1-Infectious-bronchitis-virus
ADV71773.1-Infectious-bronchitis-virus
ADV71823.1-Infectious-bronchitis-virus
AWJ64262.1-Porcine-enteric-alphaCoV
ATN23888.1-Rhinolophus-bat-CoV-HKU2
AXY04090.1-SwineAcuteDiarrheaSyndrome-related-CoV
QID98967.1-SwineAcuteDiarrheaSyndrome-related-CoV
AVM80490.1-SwineAcuteDiarrheaSyndrome-related-CoV
ABQ57223.1-Rhinolophus-bat-CoV-HKU2
ABQ57215.1-Rhinolophus-bat-CoV-HKU2
YP_001552234.1-Rhinolophus-bat-CoV-HKU2
ALA50248.1-Camel-alpha-CoV
ARQ03268.1-Porcine-epidemic-diarrhea-virus
AFU92130.1-Hipposideros-bat-CoV-HKU10
QCX35177.1-Rhinolophus-bat-CoV-HKU32
AXX83348.1-Human-CoV-OC43
AEN19363.1-Human-CoV-OC43
AGT51508.1-Human-CoV-OC43
AXX83342.1-Human-CoV-OC43
AAR01012.1-Human-CoV-OC43
YP_009555238.1-Human-CoV-OC43
AGV08377.1-MERS-CoV
AGG22540.1-Beta-CoV-2c-England-Qatar
AHX71944.1-MERS-CoV
ASU90338.1-MERS-CoV
ALR69640.1-MERS-CoV
ASU90679.1-MERS-CoV
AGN72639.1-MERS-CoV
AFY13306.1-Beta-CoV-England1
AHB33324.1-MERS-CoV
ALX27230.1-MERS-CoV
AIY60526.1-MERS-CoV
YP_009047202.1-MERS-CoV
YP_009072438.1-Bat-CoV/Zhejiang2013
YP_003858583.1-Bat-CoV-BM48-31
AAR87587.1-SARS-CoV-HKU39849
ADC35510.1-SARS-CoV-TW8
AAR87543.1-SARS-CoV-TW4
AAR86849.2-SL-CoV
NP_828849.2-SARS-CoV-TJF
AAT76146.1-SARS-CoV-Urbani
AAP13442.1-SARS-CoV-wticMB
ACZ71766.1-SARS-CoV-Frankfurt1
AAP33696.1-SARS-CoV-BJ01
AAP30028.1-SARS-CoV-CUHK-W1
AAP13566.1-SARS-CoV-A022-civet
AAV91630.1-SARS-CoV-007/2004
AAU04645.1-civet SARS-CoV-civet010
AAU04648.1-SARS-CoV-GZ02
AAS00002.1-SARS-CoV-GZ02
AKZ19085.1-Bat-SL-CoV-YNLF_34C
AIA62329.1-BtRs-YN2013
ATO98179.1-Bat-SL-CoV
ATO98191.1-Bat-SL-CoV
ARI44803.1-Bat-CoV
ATO98143.1-Bat-SL-CoV
ATO98155.1-Bat-SL-CoV
QDF43834.1-BtRs-YN2018D
QDF43824.1-BtRs-YN2018B
QDF43829.1-BtRs-YN2018C
ATO98216.1-Bat-SL-CoV
ATO98130.1-Bat-SL-CoV
AGZ48805.1-Bat-SL-CoV-RsSHC014
AGZ48830.1-Bat-SL-CoV-WIV1
AIA62319.1-BtRs-GX2013
ADE34765.1-Bat-SARS-CoV-HKU3-8
AIA62309.1-BtRs-CoV-2014
AID16715.1-Bat-SL-CoV-2013
AVP78041.1-Bat-SL-CoVZXC21
AVP78030.1-Bat-SL-CoVZC45
QHR63299.1-Bat-RaTG13
QHS34545.1-India-Kerala-State
QIZ16507.1-Turkey
QIX12193.1-Iran
QIQ22758.1-Pakistan-Gilgit
QIH45022.1-China-Beijing
QIZ15535.1-South-Africa
QIZ16545.1-Greece
QIU78705.1-Spain
QIT06961.1-USA-FL
QIS60286.1-Peru-1
QIX12146.1-France
QIC53203.1-Sweden
QHZ00398.1-US-MA1
QIT06985.1-Israel
QIT06937.1-USA-OR
QIG55993.1-Brazil
QIA98553.1-Italy
QHR84448.1-Australia-VIC01
QIB84672.1-Nepal
QHZ00378.1-SouthKorea
QHZ00357.1-China-Hangzhou
QHO60603.1-US-WA
YP_009724389.1-Wuhan-Hu-1
QIK50437.1-Viet-Nam
QII57165.1-China-Shanghai
QHZ00388.1-US-WI1
QHO62876.1-US-IL
BCA87360.1-Japan
QIS60274.1-Pakistan-kpk
QID21067.1-US-TX1
QIA98595.1-Taiwan
QHU79171.2-Finland
QHN73794.1-China-Shenzhen
BCB15089.1-Japan
QHQ82463.1-US-AZ1
QIA20042.1-China-Yunnan
QIE07450.1-China-Guangzho
QHQ71972.1-US-CA
β-CoV
SARS-CoV-2
bat-CoV
bat-SL-CoV
BtRs-Beta-CoV
SARS-CoV
MERS-CoV
Human-CoV-OC43
α-CoV
γ-CoV
δ-CoV

Fig. 1 Phylogenetic characterization of SARS-CoV-2 based on the 7096-aa polyprotein pp1ab. Phylogenetic tree demonstrating the relationship of SARS-CoV-2 to other CoVs. Phylogenetic analysis involved 121 pp1ab sequences from the subfamily *Coronavirinae*, including representatives of four genera; α-CoV, β-CoV, γ-CoV, and δ-CoV. The color codes distinguish between various groups/types of coronaviruses. The phylogenetic tree was reconstructed using the maximum likelihood method with the WAG substitution model. Bootstrap values ≥50% are shown along the branches. Scaled phylogram of this tree with branch lengths reflecting the amount of genetic change is provided in Supplementary Fig. S1.

functional uniqueness of SARS-CoV-2 using a pp1ab-based analysis. In SARS-CoV-2 genomes sampled from multiple, distinct geographic locations, we searched for amino-acid substitutions in the 7096-aa long pp1ab sequence by comparison with the closely related batSL-CoVs (RaTG13/ZC45/ZXC21) and the distantly related representatives of the β-CoV lineage (Fig. 2a–c; Supplementary Tables S1, S2). Inspection of comparative data revealed a total of 90 replacements/insertions/deletions in pp1ab of SARS-CoV-2 (Supplementary Fig. 3; Supplementary Table S3). Among these, 53 amino-acid differences were found to be fixed in pp1ab of SARS-CoV-2, as compared with batSL-CoVs (Supplementary Table S3). The Nsp3 within pp1ab appeared to be crucial to the evolutionary diversification of SARS-CoV-2, harboring 31 of the 53 fixed substitutions (Supplementary Table S3). Fixation of 31 amino-acid replacements within Nsp3 of SARS-CoV-2 was further validated through analysis of representative sequence data from seven distinct clades of SARS-CoV-2 (G, GH, GR, L, O, S, and V) reported in GISAID (https://www.gisaid.org) [18] (Supplementary Table S4; Supplementary Alignment File). Nsp3 is the



```
AAP30028.1:SARS-CoVBJ01  L A K K C L     N L N A G E   A K P L Q S           - - - V - E   - P P N T   V Q K P V D V - K   M L - R G E   P I C M D V
AVP78041.1: Bat-ZXC21    L A K N C L     N V N R G E   A D P V H S           E Q K V A E   V K P F F   E Q R Q Q A E E K   T L V K D I   P V C M E T
AVP78030.1: Bat-ZC45     L A K N C L     N V N R G E   A D P V H S           E Q K V A E   V K S S I   E Q R Q Q A E E K   T L V K D I   P V C M E T
QHR63299.1-Bat-RaTG13    L A K N C L     N V N R G E   A D P V H S           E Q K T A E   V K S S I   E Q R Q Q - V D K   T L V N D I   P V C M E T
QIQ22758.1-Pak-Gilgit    . . . H . .     . . . K . .   . . . I . .           . . . I . .   . . P F .   . . . K . - D . .   . . . S . .   . . . V . .
YP_009724389.1-China     . . . H . .     . . . K . .   . . . I . .           . . . I . .   . . P F .   . . . K . - D . .   . . . S . .   . . . V . .
QHR84448.1-Australia     . . . H . . ••• . . . K . . ••• . . . I . . ••••••••• . . . I . . ••• . . P F . ••• . . . K . - D . . ••• . . . S . . ••••••••• . . . V . .
QIA98553.1-Italy         . . . H . .     . . . K . .   . . . I . .           . . . I . .   . . P F .   . . . K . - D . .   . . . S . .   . . . V . .
QIG55993.1-Brazil        . . . H . .     . . . K . .   . . . I . .           . . . I . .   . . P F .   . . . K . - D . .   . . . S . .   . . . V . .
QIT06985.1-Israel        . . . H . .     . . . K . .   . . . I . .           . . . I . .   . . P F .   . . . K . - D . .   . . . S . .   . . . V . .
QHO60603.1-US-WA         L A K H C L     N V N K G E   A D P I H S           E Q K I A E   V K P F I   E Q R K Q - D D K   T L V S D I   P V C V E T
```

Fig. 2 Identification of sequence divergence at the polyprotein pp1ab locus in SARS-CoV-2. a Schematic of SARS-CoV-2 genome. Genomic organization of SARS-CoV-2 with numbering above the block referring to nucleotide positions. Structural proteins, including spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins, as well as nonstructural proteins (nsps) translated from ORF1ab and accessory proteins are indicated. **b** Schematic of 7096-aa replicase polyprotein pp1ab with its 15 sub-proteins (nsp1–nsp10 and nsp12–nsp16). **c** Comparative sequence analysis of the pp1ab domain for completely sequenced genomes of SARS-CoV-2 from at least 39 distinct global territories. SARS-CoV-2 sequences were compared to the corresponding homologous sequence from batSL-CoVs and SARS-CoVBJ01. Macrodomains within Nsp3 are demarcated by red arrow. *Y* axis depicts similarity scores between CoVs and the *X* axis refers to the relative residue position. Lower scores signify low sequence conservation, with trough corresponding to the least conserved regions. **d** Macrodomains sequences encoding Mac-1, Mac-2, and Mac-3 proteins exhibiting exceptional divergence in SARS-CoV-2 relative to bat-RaTG13/ZC45/ZXC21 and SARS-CoVBJ01. Green, red, and blue circles, respectively, differentiate between Mac-1, Mac-2, and Mac-2 specific substitutions.

largest multi-domain protein produced by CoVs, playing many important roles in the viral life cycle. In particular, macrodomains encoded within Nsp3 of CoVs have been demonstrated as critical in counteracting the host innate immune response [42].

Given their roles in virulence and pathogenesis, macrodomains of SARS-CoV-2 were subjected to further scrutiny through sequence, structural, and evolutionary analysis. Macrodomains located in Nsp3 were found to be particularly enriched with fixed amino-acid replacements specific

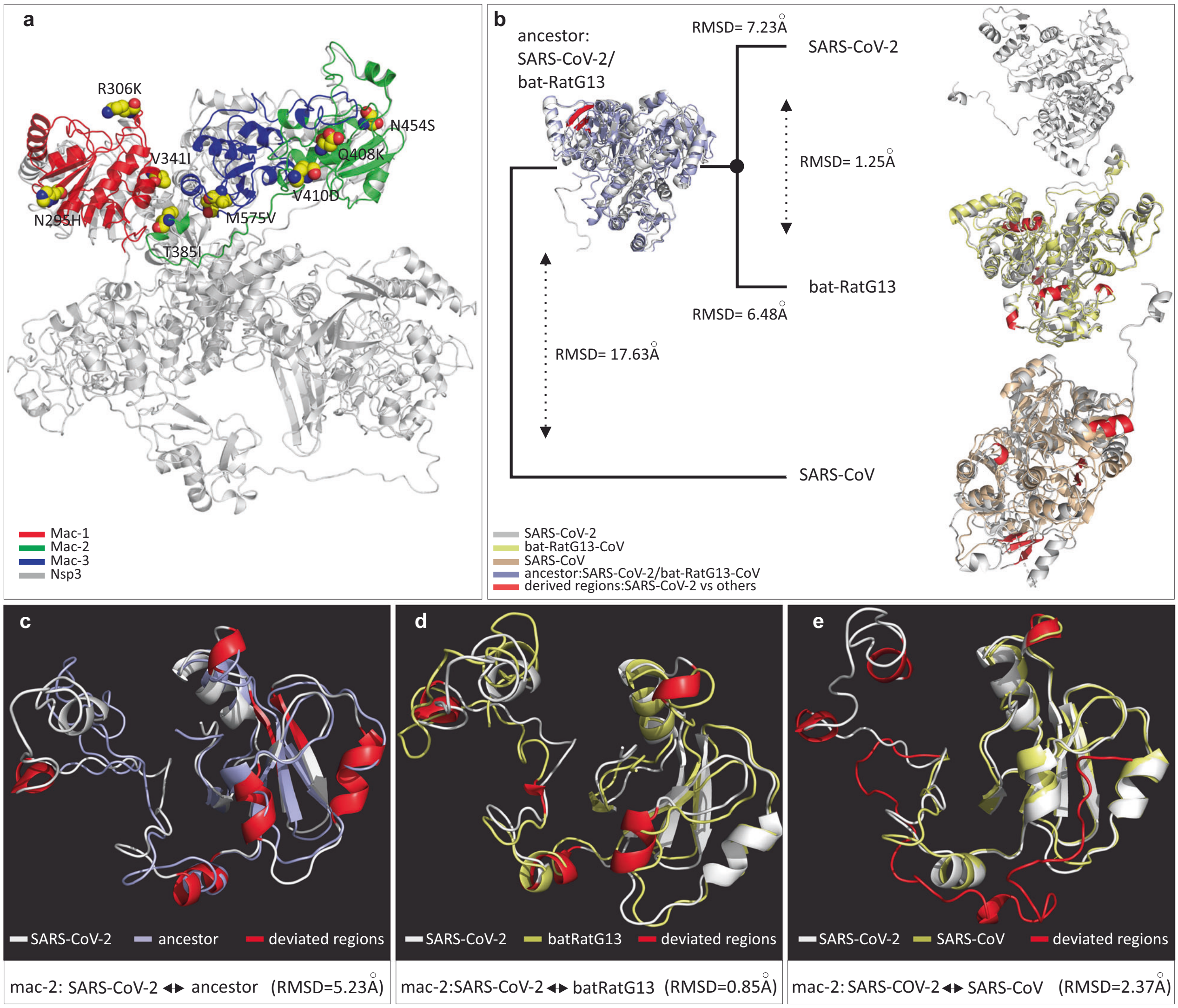


Fig. 3 Protein structural analysis of macrodomains examining the effects of specific fixed substitutions within Nsp3 of SARS-CoV-2. **a** Protein structure of SARS-CoV-2 Nsp3 depicting types and locations of fixed amino-acid substitutions within distinct macrodomains. The fixed amino-acid replacements are shown as spheres and labeled with the amino-acid position in Nsp3 protein. **b** Comparison of the three-dimensional (3D) conformations of the macrodomains (Mac-1, Mac-2, and Mac-3) within Nsp3. SARS-CoV-2 (YP_009725299.1) macrodomains were also compared to the corresponding homologous protein regions of bat-RatG13 (QHR63299.1), SARS-CoVBJ01 (AAP30028.1), and the predicted ancestral macrodomain of SARS-CoV-2 and bat-RaTG13/ZC45/ZXC21. Structural deviations in terms of backbone torsion angles (Φ°, Ψ°) are represented in red color and were examined by RMSD (root mean square deviation) values. **c–e** Close-up of 3D conformations of SARS-CoV-2 Mac-2 with corresponding homologous domains from predicted ancestor (aforementioned; left panel), bat-RatG13 (middle panel), and SARS-CoV (AAP30028.1; right panel). Comparisons of 3D conformations for Mac-1 is provided in Fig. 4 and sequence secondary structural level details for comparisons in **b–e** are given in Supplementary Tables S6, S7.

to SARS-CoV-2 (Fig. 2d). In total, eight substitutions were fixed in the macrodomains of SARS-CoV-2 (Supplementary Table S5). These eight substitutions, though divergent from corresponding homologous positions in closely related bats (RaTG13/ZC45/ZXC21), are otherwise exceptionally constrained; not a single amino-acid difference was noted among the 39 SARS-CoV-2 genomes at these sites (Supplementary Fig. S4). Conceivably, these substitutions are the consequence of accelerated rates of sequence evolution, which may have been driven by positive Darwinian selection after the divergence of SARS-CoV-2 and batSL-CoVs and prior to its first reported outbreak. Among the identified set of macrodomain-specific fixed substitutions, three were found to reside within macrodomain 1 (Mac-1), four in macrodomain 2 (Mac-2), whereas only a single amino-acid substitution was found to be fixed in macrodomain 3 (Mac-3) of SARS-CoV-2 (Fig. 3a).

Next, we further determined the impact of the fixed amino-acid substitutions identified within the macrodomains of Nsp3 of SARS-CoV-2 by predicting the thermodynamic state function. For this we determined the ΔΔG (Gibbs free energy) of a substitution in SARS-CoV-2 relative to the ancestral protein ($\Delta\Delta G_{\mathrm{ances}\rightarrow\mathrm{CoV2}}$) [27]. All of the eight fixed replacements appeared to have significant destabilizing effects on protein structure ($\Delta\Delta G < 0.0$) (Supplementary Table S5). Since mutations that modulate enzymatic functions or ligand binding are often destabilizing [43], the thermodynamic effects of the substitutions unique to SARS-CoV-2 macrodomains support their adaptive significance. In addition, examination of physicochemical property changes resulting from amino-acid substitutions in the SARS-CoV-2 macrodomains predict that, with respect to their effects on protein structure and function, all fixed replacements in macrodomains of SARS-CoV-2 are of radical type, implying their biological significance (Supplementary Table S5).

The globular Mac-1 of Nsp3 contains a conserved cleft that binds ADP-ribose (ADPr). Mac-1 was found to possess hydrolase activity that removes ADPr from target proteins, a biochemical feature of SARS-CoV-2 that is considered essential in counteracting the host antiviral response of ADP-ribosylation [42]. Reducing the capacity for Mac-1 to remove ADPr, especially in CoVs, results in an attenuation of virulence and a greater sensitivity to host innate immune responses [44–46]. Furthermore, the Macrodomain 2 and Macrodomain 3 (Mac-2 and 3) of SARS-CoVs are known to be indispensable for its replication or transcription, as these macrodomains bind nucleic acids, with a preference for purine-rich RNA sequences, such as G-rich stretches [47]. mRNAs for host antiviral responses and apoptotic signaling harbor long poly(G) stretches at their 3′ untranslated regions, and thus are prime targets for Mac-2 and 3 mediated disruption of host immunity [48, 49]. Furthermore, it has been demonstrated that Mac-2 and 3, together with the papain-like protease domain (PL2pro) of Nsp3, interact with RCHY1(E3 ubiquitin ligase) and intensify RCHY1-mediated ubiquitination, which consequently induces p53 degradation. Hence, human SARS-CoVs via their Mac-2 and 3 domains downregulate p53, a major determinant of antiviral innate immunity, thus leading to delayed activation of p53-targeted immunity genes [49]. Intriguingly, Mac-2 and 3 are specifically present in Nsp3 of SARS-CoVs and highly related viruses (batSL-CoVs) known to cause high levels of pathogenicity in humans, but are not present in CoVs that cause mild infections [42].

Herein, we sought to evaluate the functional significance of fixed amino-acid replacements specific to SARS-CoV-2 by modeling the macrodomains (Mac-1, 2, and 3) of SARS-CoV-2 (YP_009725299.1), bat-RaTG13 (QHR63299.1), and SARS-CoV-BJ01 (AAP30028.1). Furthermore, the ancestral macrodomain protein sequence of SARS-CoV-2 and bat-RaTG13/ZC45/ZXC21 was also predicted and modeled. The 3D superimposition of macrodomain structures revealed that during the course of evolution, SARS-CoV-2 macrodomains have had significant transitions in various secondary structural elements (SSEs) (Fig. 3b). A continuous transition from loops to core SSEs was observed (Supplementary Table S6). For instance, the macrodomains of SARS-CoV-2 contain about 62% of their residues in SSEs (Supplementary Table S6). In contrast, the macrodomains of SARS-CoV had about 41% of their total residues occurring in loops and the remaining 59% were in SSEs, such as α-helices and β-sheets (Supplementary Table S6). Furthermore, multiple substitutions scattered across the SARS-CoV-2 macrodomains were found to reposition specific protein regions within Mac-1 and Mac-2, in three-dimensional space (Figs. 3c–e, 4; Supplementary Table S7). In particular, there were drastic changes in the conformation of Mac-2 of SARS-CoV-2 (Fig. 3c–e). Structural evolution of Mac-2 involves considerable alterations in sequence, length, and conformation of core SSEs, implicating a functional relevance of these substitutions (Supplementary Table S7).

To further connect protein conformational changes with functions of macrodomains, we used a molecular docking approach to determine the binding affinity of ADPr to Mac-1. Structural comparisons revealed that there is considerable divergence in ADPr binding between SARS-CoV-2 and other CoVs (Fig. 5b–d). ADPr binding to the Mac-1 domain of SARS-CoV-2 (−9.46 kcal/mol) was found to be more efficient than in the human SARS-CoV (−8.59 kcal/mol) (Supplementary Table S8). Intriguingly, the binding affinity of SARS-CoV-2 Mac-1 for ADPr was comparable to that of MERS-CoV Mac-1 (−9.70 kcal/mol). This suggest that SARS-CoV-2 may evade host antiviral ADPr activity similar to that of the highly pathogenic MERS-CoV [50].

To gain insights into the molecular mechanisms behind the similar ADPr binding affinities of SARS-CoV-2 and MERS-CoV macrodomains, we investigated their binding clefts. In MERS-CoV, Asp-20 within the α1-helix of Mac-1 has been shown to be critical for binding specificity; its side chain forms a direct contact with ADPr through hydrogen bonding with the N-6 atom of the pyrimidine ring in the adenine moiety [51] (Fig. 5a, f, g). This residue is known to be conserved among macrodomains of CoVs. It determines the degrees of curvature at the adenine moieties within binding pocket, and hence the binding affinity of ADPr [51] (Fig. 5a). Previous reports that superposition the MERS-CoV macrodomains showed that the same oxygen atom on the Asp-20 side chain forms two hydrogen bonds; one with the N-6 atom in a pyrimidine ring of ADPr and the other with nitrogen in the Ile-22 backbone amide in the α1-helix. This results in the displacement of the Asp-20 side chain into the adenine cavity, which strengthens ADPr binding to effectively limit host antiviral ADPr activity [51]. In contrast, the equivalent residue of SARS-CoV, Asp-23 forms a

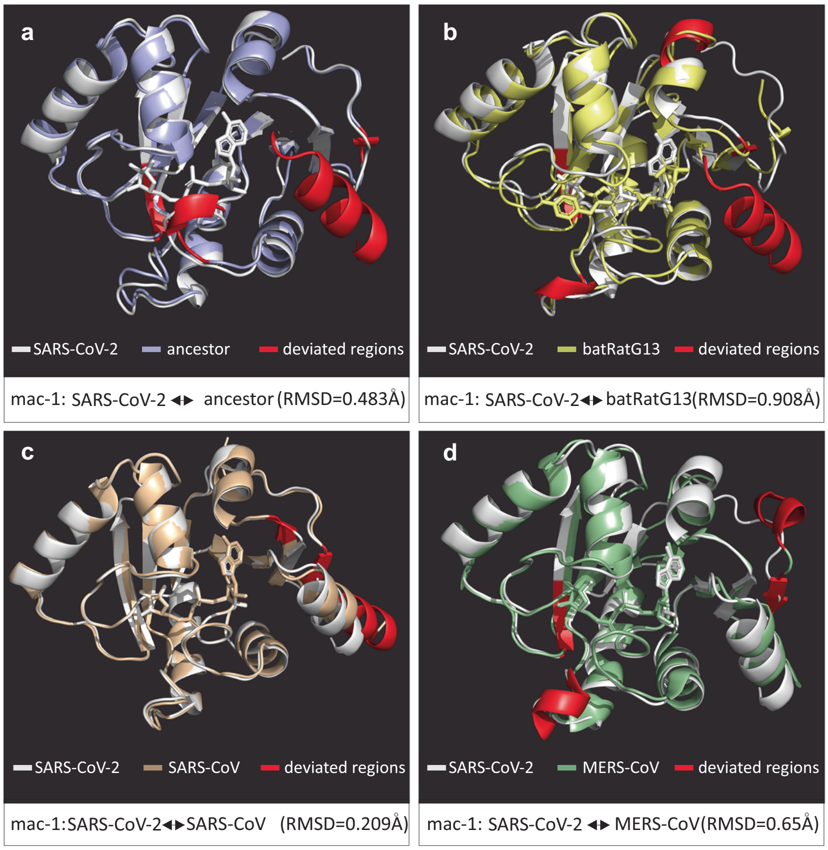

a
SARS-CoV-2
ancestor
deviated regions
mac-1: SARS-CoV-2 ◂▸ ancestor (RMSD=0.483Å)
b
SARS-CoV-2
batRatG13
deviated regions
mac-1: SARS-CoV-2◂▸batRatG13(RMSD=0.908Å)
c
SARS-CoV-2
SARS-CoV
deviated regions
mac-1:SARS-CoV-2◂▸SARS-CoV (RMSD=0.209Å)
d
SARS-CoV-2
MERS-CoV
deviated regions
mac-1: SARS-CoV-2 ◂▸ MERS-CoV(RMSD=0.65Å)

Fig. 4 Comparisons of 3D conformations of SARS-CoV-2 macrodomain 1. Comparisons of 3D conformations of SARS-CoV-2 macrodomain 1 with corresponding homologous domains from (**a**) predicted ancestral macrodomain 1 (of SARS-CoV-2 and bat-RaTG13/ZC45/ZXC21), (**b**) bat-RatG13 (QHR63299.1), (**c**) SARS-CoV (PDB entry: 2FAV), and (**d**) MERS-CoV (PDB entry: 5HOL). Descriptions of color codes are given in each panel. Deviated residues in terms of backbone torsion angles (Φ°, Ψ°) are shown in red color. Structural deviations were examined by RMSD values. Note: Primary sequence and secondary structural level details for comparisons in **a**–**d** are given in Supplementary Tables S6, S7.

only single hydrogen bond with ADPr via an oxygen atom in its side chain and binds to nitrogen atoms in Val-25 and Lys-26 of the α1-helix via another hydrogen bond [51]. Intriguingly, inspection of equivalent hydrogen bonding patterns in SARS-CoV-2 revealed a closer resemblance to MERS-CoV than to SARS-CoV (Fig. 5e, f). For example, in SARS-CoV-2, the oxygen atom on Asp22 side chain forms two hydrogen bonds, one with the N-6 atom in a pyrimidine ring of ADPr and the other with nitrogen in the Val-24 of the α1-helix, which is similar to MERS-CoV, and, accordingly, may displace the Asp22 side chain into the adenine cavity (Fig. 5f). Furthermore, studies of MERS-CoV provide evidence that the stabilization of ADPr in the binding pocket of Mac-1 is mainly determined by hydrogen-bond strength [51]. In the MERS-CoV the hydrogen-bond lengths formed by the Asp-20 side chain with ADPr and Ile-22 are 2.9 and 3.0 Å, respectively (Fig. 5f). This is comparable to the SARS-CoV-2 hydrogen-bond lengths formed by the Asp-23 side chain with ADPr and Val-24, which for both is 2.9 Å (Fig. 5e, f). Conceivably, the comparable affinities of MERS-CoV and SARS-CoV-2 macrodomains for ADPr may partly result from similarities in the their hydrogen bonding patterns, hydrogen-bond lengths/strength, and the positioning of the side chains of Asp-20/Asp-22 with respect to cleft that holds the adenine moiety. Thus, it appears that the biophysical and structural aspects of ADPr binding site evolution may provide SARS-CoV-2 and MERS-CoV with extraordinary adaptive abilities, which enable these viruses to evade host innate defense pathways. To further validate these findings, we analyzed the druggability scores of ADPr binding pockets of Mac-1 of SARS-CoV, MERS-CoV, and SARS-CoV-2 by using DoGSiteScorer web server [41]. Binding clefts of MERS-CoV and SARS-CoV-2 depict high druggability scores as compared to SARS-CoV (Supplementary Table S9). Higher druggability scores are considered to reflect a greater ability of protein pocket to bind to its target ligand [41]. Thus, druggability assessment suggests that macrodomains of MERS-CoV and SARS-CoV-2 are more druggable and hence provide better binding cavities for conformational optimization of ADPr as compared to SARS-CoV.

Previously, it was suggested that the differential binding affinity of Mac-1 for ADPr may result from different amino-acid compositions in the α1-helix [51]. Here, we showed that, though the corresponding residues within the α1-helices of SARS-CoV-2 and MERS-CoV are highly divergent, their binding affinities for ADPr are similar (Fig. 5a; Supplementary Table S8). Therefore, it is possible that the evolved increase in binding affinity of SARS-CoV-2 Mac-1 for ADPr may have been facilitated by epistatic effects of three fixed amino-acid substitutions located near the binding

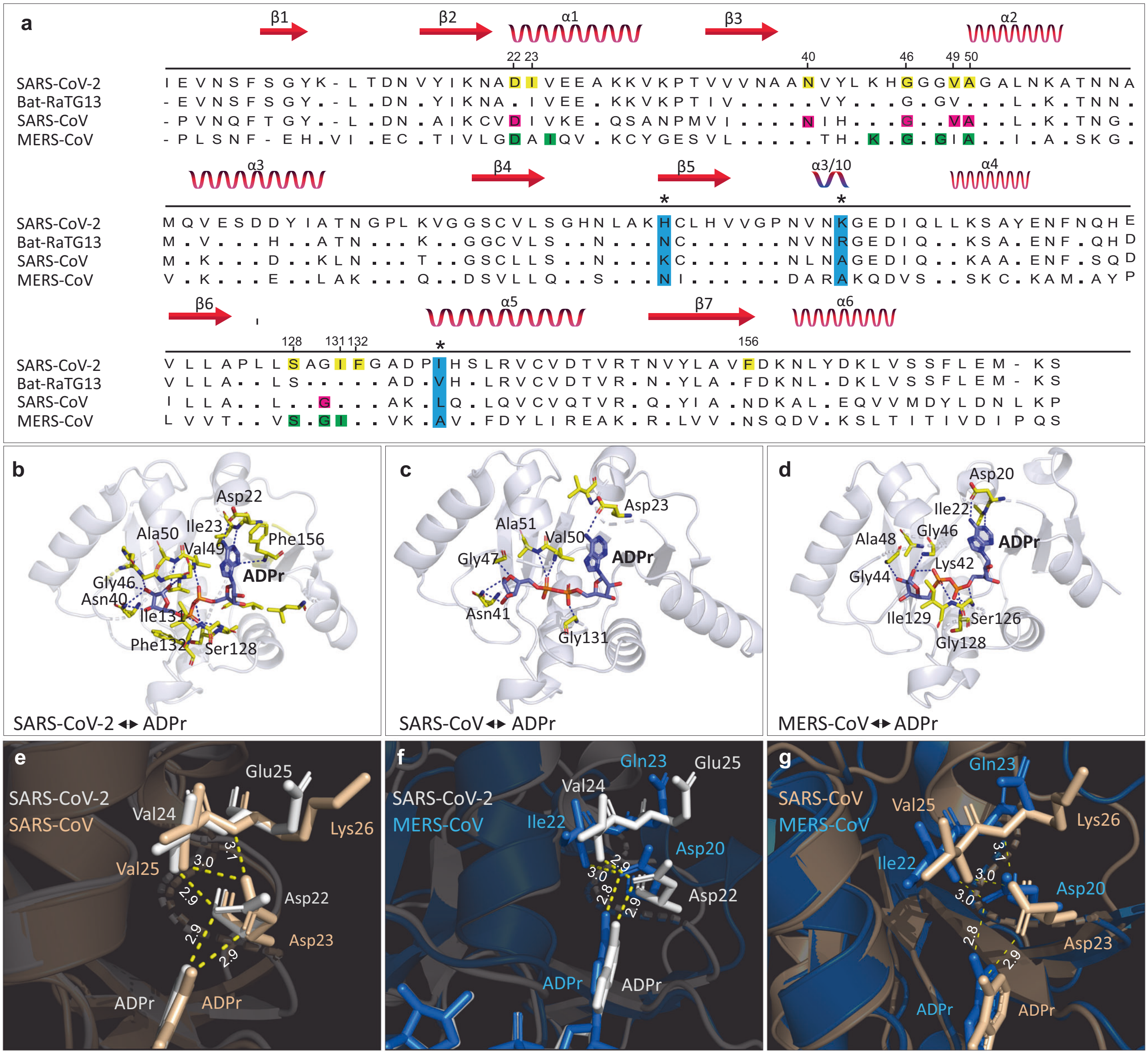

a
β1
β2
α1
β3
α2
22 23
40
46
49 50
SARS-CoV-2 I E V N S F S G Y K - L T D N V Y I K N A D I V E E A K K V K P T V V V N A A N V Y L K H G G G V A G A L N K A T N N A
Bat-RaTG13 - E V N S F S G Y . - L . D N . Y I K N A . I V E E . K K V K P T I V . . . . . V Y . . . G . G V . . . L . K . T N N .
SARS-CoV - P V N Q F T G Y . - L . D N . A I K C V D I V K E . Q S A N P M V I . . . . N I H . . . G . . V A . . L . K . T N G .
MERS-CoV - P L S N F - E H . V I . E C . T I V L G D A I Q V . K C Y G E S V L . . . . . T H . K . G . G I A . . I . A . S K G .
α3
β4
β5
α3/10
α4
*
*
SARS-CoV-2 M Q V E S D D Y I A T N G P L K V G G S C V L S G H N L A K H C L H V V G P N V N K G E D I Q L L K S A Y E N F N Q H E
Bat-RaTG13 M . V . . . H . . A T N . . . K . . G G C V L S . . N . . . N C . . . . . . N V N R G E D I Q . . K S A . E N F . Q H D
SARS-CoV M . K . . . D . . K L N . . . T . . G S C L L S . . N . . . K C . . . . N L N A G E D I Q . . K A A . E N F . S Q D
MERS-CoV V . K . . . E . . L A K . . . Q . . D S V L L Q . . S . . . N I . . . . D A R A K Q D V S . . S K C . K A M . A Y P
β6
α5
β7
α6
128
131 132
*
156
SARS-CoV-2 V L L A P L L S A G I F G A D P I H S L R V C V D T V R T N V Y L A V F D K N L Y D K L V S S F L E M - K S
Bat-RaTG13 V L L A . . L S . . . . . A D . V H . L R V C V D T V R . N . Y L A . F D K N L . D K L V S S F L E M - K S
SARS-CoV I L L A . . L . . G . . . A K . L Q . L Q V C V Q T V R . Q . Y I A . N D K A L . E Q V V M D Y L D N L K P
MERS-CoV L V V T . . V S . G I . . V K . A V . F D Y L I R E A K . R . L V V . N S Q D V . K S L T I T I V D I P Q S
b
Asp22
Ala50
Ile23
Val49
Phe156
ADPr
Gly46
Asn40
Ile131
Phe132
Ser128
SARS-CoV-2 ◂▸ ADPr
c
Asp23
Ala51
Val50
Gly47
ADPr
Asn41
Gly131
SARS-CoV ◂▸ ADPr
d
Asp20
Ile22
Ala48
Gly46
ADPr
Lys42
Gly44
Ile129
Ser126
Gly128
MERS-CoV ◂▸ ADPr
e
Glu25
SARS-CoV-2
SARS-CoV
Val24
Lys26
Val25
3.1
3.0
2.9
Asp22
Asp23
2.9
2.9
ADPr
ADPr
f
Gln23
Glu25
SARS-CoV-2
MERS-CoV
Val24
Ile22
Asp20
3.0
2.9
2.8
2.9
Asp22
ADPr
ADPr
g
Gln23
SARS-CoV
MERS-CoV
Val25
Lys26
Ile22
3.1
3.0
3.0
Asp20
Asp23
2.8
2.9
ADPr
ADPr

Fig. 5 Sequence and structural comparison of Mac-1 protein in SARS-CoV-2 to other CoVs and Mac-1 capacity to bind ADPr. a Comparison of Mac-1 sequence of SARS-CoV-2 (YP_009725299.1) to that of MERS-CoV (PDB entry: 5HOL), SARS-CoV (PDB entry: 2FAV), and bat-RatG13 (QHR63299.1). Secondary structure elements are depicted in red at the top of the alignment. Alignment positions with fixed substitutions within Mac-1 that are specific to SARS-CoV-2 are highlighted in light blue columns with asterisk (*) symbol. Amino-acid residues that form a hydrogen bond with ADP-ribose (ADPr) are highlighted as yellow, pink, and green for SARS-CoV-2, SARS-CoV, and MERS-CoV, respectively. **b–d** Structures of the Mac-1 complex with ADPr. In each case, amino-acid residues that form a hydrogen bond with ADPr molecule are shown. **e–g** Comparison of interactions with ADPr in the Mac-1 adenine pocket of SARS-CoV-2, MERS-CoV, and SARS-CoV. Hydrogen bonds are shown as yellow dashed lines and bond lengths are given in Å units.

cleft, but that do not contact the ADPr via direct hydrogen bonding (Fig. 5a). Evolution by protein conformational epistasis may play a significant role in differential binding affinities of Mac-1 for ADPr [43]. Interestingly, differential binding affinities of Mac-1 for ADPr have been associated with the differences in the pathogenicity of coronaviruses [45, 51]. Thus, the comparable affinities of SARS-CoV-2 and MERS-CoV macrodomains for ADPr fits well with the notion that current mortality rates grossly underestimate the threat posed by COVID-19 [6].

This study has revealed functionally unique amino-acid replacements within the macrodomains of Nsp3, that are likely to maximize SARS-CoV-2 activity against human innate immune responses. The precise residue-level structural information in our study may benefit the design of anti-SARS-CoV-2 drug treatments. Macrodomains have previously been shown to be potent drug targets. For instance, an AlphaScreen based assay has recently identified a small molecule inhibitor, GeA-69 (a carbazole-based compound) that targets the macrodomain of human PARP14 (poly-ADP-ribose polymerase 14), a pro-survival protein associated with human inflammatory diseases and various types of cancers [52]. In addition, the combination of structure-based virtual screening and molecular dynamics simulation approaches have been successful in identifying potential inhibitors targeting viral macrodomains [53].

Acknowledgements The authors thank Prof. Neil H. Shubin (The University of Chicago) for useful comments and extensive guidance.

Funding This work was supported by National Key Research and Development Program of China [2016YFE0206600 to YB]; The 13th 5-year Informatization Plan of Chinese Academy of Sciences [XXH13505-05 to YB]; The Open Biodiversity and Health Big Data Initiative of IUBS [to YB]. VL is supported by grants from the Department of Defense (W81XWH1810512), the National Institute of Neurological Disorders and Stroke (1R21NS112614-01 and 1R01NS113894-01A1), and a Gibby & Friends vs. Parky Award. DQW is supported by the grants from the Key Research Area Grant 2016YFA0501703 of the Ministry of Science and Technology of China, the National Natural Science Foundation of China (Contract nos. 61832019, 61503244), the Science and Technology Commission of Shanghai Municipality (Grant: 19430750600), the Natural Science Foundation of Henan Province (162300410060) and Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2017ZD14). The computations were partially performed at the Pengcheng Lab. And the Center for High-Performance Computing, Shanghai Jiao Tong University.

Author contributions AAA and YB conceived the project. AAA, YB, DQW, and VL designed the experiments. AAA, IH, NP, AK, SS, and HS performed the experiments. AAA, VL, YB, DQW, IH, NP, and AH analyzed the data. AAA, VL, YB, DQW, IH, NP, and AK, wrote the paper.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579:270–3.
2. Luk HKH, Li X, Fung J, Lau SKP, Woo PCY. Molecular epidemiology, evolution and phylogeny of SARS coronavirus. Infect Genet Evol. 2019;71:21–30.
3. Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, et al. Bats are natural reservoirs of SARS-like coronaviruses. Science. 2005; 310:676–9.
4. Perlman S, Netland J. Coronaviruses post-SARS: update on replication and pathogenesis. Nat Rev Microbiol. 2009;7:439–50.
5. Chen Y, Liu Q, Guo D. Emerging coronaviruses: genome structure, replication, and pathogenesis. J Med Virol. 2020;92:418–23.
6. Baud D, Qi X, Nielsen-Saines K, Musso D, Pomar L, Favre G. Real estimates of mortality following COVID-19 infection. Lancet Infect Dis. 2020. https://doi.org/10.1016/S1473-3099(20)30195-X.
7. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science. 2020;368:489–93.
8. Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, et al. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. Cell Host Microbe. 2020;27:325–8.
9. Pang J, Wang MX, Ang IYH, Tan SHX, Lewis RF, Chen JI-P, et al. Potential rapid diagnostics, vaccine and therapeutics for 2019 novel coronavirus (2019-nCoV): a systematic review. J Clin Med. 2020;9:623.
10. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. Proc Natl Acad Sci USA. 2020;117:9241–3.
11. Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, et al. Structural basis of receptor recognition by SARS-CoV-2. Nature. 2020;581: 221–4.
12. Soltis DE, Albert VA, Savolainen V, Hilu K, Qiu Y-L, Chase MW, et al. Genome-scale data, angiosperm relationships, and "ending incongruence": a cautionary tale in phylogenetics. Trends Plant Sci. 2004;9:477–83.
13. Rannala B, Yang Z. Phylogenetic inference using whole genomes. Annu Rev Genom Hum Genet. 2008;9:217–31.
14. Rokas A, Williams BL, King N, Carroll SB. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature. 2003;425:798–804.
15. Sayers EW, Beck J, Brister JR, Bolton EE, Canese K, Comeau DC, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2020;48:D9–16.
16. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 2007;35:D61–5.
17. Zhao W-M, Song S-H, Chen M-L, Zou D, Ma L-N, Ma Y-K, et al. The 2019 novel coronavirus resource. Yi Chuan Hered. 2020; 42:212–21.
18. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Chall. 2017;1:33–46.
19. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22:4673–80.
20. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28:2731–9.

21. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol. 2001;18:691–9.
22. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4:406–25.
23. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Bioinformatics 1992;8:275–82.
24. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30:3059–66.
25. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7:539–9.
26. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet 2000, 16: 276–7.
27. Pandurangan AP, Ochoa-Montaño B, Ascher DB, Blundell TL. SDM: a server for predicting effects of mutations on protein stability. Nucleic Acids Res. 2017;45:W229–35.
28. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci USA. 1992;89:10915.
29. Chou K-C, Shen H-B. A new method for predicting the subcellular localization of eukaryotic proteins with both single and multiple sites: Euk-mPLoc 2.0. PLoS ONE. 2010;5:e9931.
30. Löytynoja A, Goldman N. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. Science. 2008;320:1632–5.
31. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.
32. Krieger E, Nabuurs SB, Vriend G. Homology modeling. Methods Biochem Anal. 2003;44:509–24.
33. Webb B, Sali A. Comparative protein structure modeling using MODELLER. Curr Protoc Bioinform. 2016;54:5.6.1–37.
34. Rose PW, Prlić A, Altunkaya A, Bi C, Bradley AR, Christie CH, et al. The RCSB protein data bank: integrative view of protein, gene and 3D structural information. Nucleic Acids Res. 2017;45: D271–81.
35. DeLano WL. Pymol: an open-source molecular graphics tool. CCP4 Newsl Protein Crystallogr. 2002;40:82–92.
36. Hooft RW, Sander C, Vriend G. Objectively judging the quality of a protein structure from a Ramachandran plot. Bioinformatics. 1997;13:425–30.
37. Maiorov VN, Crippen GM. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. J Mol Biol. 1994;235:625–34.
38. Buchan DW, Jones DT. The PSIPRED protein analysis workbench: 20 years on. Nucleic Acids Res. 2019;47:W402–7.
39. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem. 2009;30:2785–91.
40. Clark AJ, Tiwary P, Borrelli K, Feng S, Miller EB, Abel R, et al. Prediction of protein–ligand binding poses via a combination of induced fit docking and metadynamics simulations. J Chem Theory Comput. 2016;12:2990–8.
41. Volkamer A, Kuhn D, Rippmann F, Rarey M. DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. Bioinformatics. 2012;28:2074–5.
42. Fehr AR, Jankevicius G, Ahel I, Perlman S. Viral macrodomains: unique mediators of viral replication and pathogenesis. Trends Microbiol. 2018;26:598–610.
43. Tokuriki N, Stricher F, Serrano L, Tawfik DS. How protein stability and new functions trade off. PLOS Comput Biol. 2008;4: e1000002.
44. Fehr AR, Channappanavar R, Jankevicius G, Fett C, Zhao J, Athmer J, et al. The conserved coronavirus macrodomain promotes virulence and suppresses the innate immune response during severe acute respiratory syndrome coronavirus infection. mBio. 2016;7:e01721–16. https://doi.org/10.1128/mbio.01721-16.
45. Kuri T, Eriksson KK, Putics A, Züst R, Snijder EJ, Davidson AD, et al. The ADP-ribose-1”-monophosphatase domains of severe acute respiratory syndrome coronavirus and human coronavirus 229E mediate resistance to antiviral interferon responses. J Gen Virol. 2011;92:1899–905.
46. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. Methods Mol Biol. 2015;1282:1–23.
47. Kusov Y, Tan J, Alvarez E, Enjuanes L, Hilgenfeld R. A G-quadruplex-binding macrodomain within the “SARS-unique domain” is essential for the activity of the SARS-coronavirus replication–transcription complex. Virology. 2015;484:313–22.
48. Tan J, Kusov Y, Mutschall D, Tech S, Nagarajan K, Hilgenfeld R, et al. The “SARS-unique domain” (SUD) of SARS coronavirus is an oligo(G)-binding protein. Biochem Biophys Res Commun. 2007;364:877–82.
49. Ma-Lauer Y, Carbajo-Lozoya J, Hein MY, Müller MA, Deng W, Lei J, et al. p53 down-regulates SARS coronavirus replication and is targeted by the SARS-unique domain and PLpro via E3 ubiquitin ligase RCHY1. Proc Natl Acad Sci USA. 2016;113:E5192–201.
50. Frick DN, Virdi RS, Vuksanovic N, Dahal N, Silvaggi NR. Molecular Basis for ADP-ribose binding to the macro-X domain of SARS-CoV-2 Nsp3. bioRxiv. 2020;59:2608–15.
51. Cho C-C, Lin M-H, Chuang C-Y, Hsu C-H. Macro domain from middle east respiratory syndrome coronavirus (MERS-CoV) is an efficient ADP-ribose binding module: CRYSTAL STRUCTURE AND BIOCHEMICAL STUDIES. J Biol Chem. 2016;291: 4894–902.
52. Schuller AP, Wu CC-C, Dever TE, Buskirk AR, Green R. eIF5A functions globally in translation elongation and termination. Mol Cell. 2017;66:194–205.e5.
53. Nguyen PTV, Yu H, Keller PA. Discovery of in silico hits targeting the nsP3 macro domain of chikungunya virus. J Mol Model. 2014;20:2216.