Variable Macro X Domain of SARS-CoV-2 Retains the Ability to Bind ADP-ribose

David N. Frick,[1]* Rajdeep S. Virdi,[1] Nemanja Vuksanovic,[1] Narayan Dahal,[2] & Nicholas R Silvaggi[1]

Departments of Chemistry & Biochemistry,[1] and Physics,[2] The University of Wisconsin- Milwaukee, Milwaukee, WI 53217

KEYWORDS: COVID-19, Antiviral Drug target, Severe Acute Respiratory Syndrome, Coronavirus.

Supporting Information Placeholder

ABSTRACT: The virus that causes COVID-19, SARS-CoV-2, has a large RNA genome that encodes numerous proteins that might be targets for antiviral drugs. Some of these proteins, such as the RNA-dependent RNA polymers, helicase and main protease, are well conserved between SARS-CoV-2 and the original SARS virus, but several others are not. This study examines one of the most novel proteins encoded by SARS-CoV-2, a macrodomain of nonstructural protein 3 (nsp3). Although 26% of the amino acids in this SARS-CoV-2 macrodomain differ from those seen in other coronaviruses, the protein retains the ability to bind ADP-ribose, which is an important characteristic of beta coronaviruses, and potential therapeutic target.

The development of antivirals targeting Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), the causative agent of the present COVID-19 pandemic,[1] will most likely focus on viral proteins and enzymes needed for replication.[2] Like other coronaviruses, SARS-CoV-2 has a large positive sense (+)RNA genome over 30,000 nucleotides long with several open reading frames. Most of the proteins that form the viral replicase are encoded by the "rep 1ab" reading frame, which codes for a 7,096 amino acid-long polyprotein that is ultimately processed into at least 15 functional peptides, five of which are only produced by a translational frameshift event occurring after nsp10 (Fig. 1). Parts of the SARS-CoV-2 rep 1ab polyprotein are very similar to the rep 1ab protein of the coronavirus that caused the SARS epidemic in 2003 (which will be referred to here as SARS-CoV-1), suggesting the that drugs targeting the SARS-CoV-1 nsp5-14 might be effective against SARS-CoV-2. However, some portions of the SARS rep 1ab polyproteins are quite different.

In contrast to the well-conserved SARS nsp5 protease, nsp12 polymerase and nsp13 helicase enzymes, significantly more differences exist between the nsp3 proteins encoded by SARS-CoV-1 and SARS-CoV-2. The most variation occurs in a domain of Nsp3 domain suspected to bind ADP-ribose, that will be referred to here as the "macro X" domain, to differentiate it from the two downstream "SARS unique macrodomains (SUDs)," which do not bind ADP-ribose.[3] The macro X domain of SARS-CoV-1 is also able to catalyze the hydrolysis of ADP-ribose 1'' phosphate, albeit at a slow rate.[4]

All the above differences preclude the use of the SARS-CoV-1 macro X domain structures as scaffolds to design compounds that might target this nsp3 region in SARS CoV-2, especially in light of the observation that the same nsp3 domain from gamma coronaviruses folds with a closed binding site and does not bind ADP-ribose *in vitro*.[5] Because ADP-ribose binding is a property that could be used to identify possible antiviral agents, the ability of the SARS-CoV-2 macro X domain to bind ADP-ribose was examined using a recombinant purified protein and isothermal titration calorimetry (ITC).

RESULTS AND DISCUSSION

Hypervariability in the nsp3 Macro X domain— The structures of most of the soluble portions of the SARS-CoV-1 nsp proteins have been examined at atomic resolution to help understand coronavirus replication and facilitate antiviral drug discovery. The amino acid sequences of each of these proteins were compared with the homologous regions of the rep 1ab protein encoded by SARS-CoV-2 (GenPept Accession YP_009724389). The most similar proteins were the RNA helicases (nsp13) which are identical in all but one of their 603 amino acids: a conservative Val to Ile substitution near their C-termini. The RNA-dependent RNA polymerases (nsp12) are also well-conserved, sharing all but 34 of 955 amino acids. The primary protease that cleaves the polyprotein (nsp5) is also similar in SARS-CoV-1 and SARS-CoV-2, with only 13 amino acids that differ among 306 (4.2% different) (Fig. 1).

At the other end of the spectrum are the nsp3 proteins, which are notably more different in the two SARS viruses. Nsp3 is a large multidomain membrane-bound protein,[6] and its clearest role in viral replication is cleaving the rep polyprotein. Over 17% of the amino acids in

the nsp3 protease domain differ between SARS-CoV-1 and SARS-CoV-2. Other parts of nsp3 are even more variable, like the macrodomains that lie N-terminal to the nsp3 protease domain. Macrodomains consist of four helices that surround a mixed beta sheet. A ligand-binding pocket that typically binds ADP-ribose or related compounds lies between the helices and the sheet.[7] SARS-CoV-1 has three macrodomains in tandem, but only the first binds ADP ribose. The amino acid sequences of this macro X domain differ by 26% between SARS-CoV-1 and SARS-CoV-2 (Fig. 1).

Many of the 47 variant residues in the 180 amino acid long SARS macro X domain are clustered near its N-terminus in a region that is particularly variable in the three *Coronaviridae* genera (Fig. 2A). Macro X domains from coronaviruses that cause the common cold (alpha coronaviruses)[5] and the beta coronaviruses like SARS-CoV-1 and Middle East respiratory syndrome coronavirus (MERS-CoV)[8] all bind ADP ribose. However, the same domain in gamma coronaviruses does not bind ADP-ribose.[5]

When atomic models of the SARS-CoV-2 macro X domain were generated with the MODELLER[24] plugin for UCSF Chimera (v. 1.14)[23] using the structure of the SARS-CoV-1 macro X domain (PDB file 2fav)[9] as a template, most of the two protein backbones superimposed, but the N-termini did not (Fig. 2B, C). In fact, the SARS-CoV-2 model with the lowest zDOPE score (-0.61) folds with the N-terminus blocking the ADP ribose binding site (Fig. 2C), which is reminiscent of what is seen in in the gamma coronaviruses (PDB file 3eke).[5] However, in other models, the protein folds in a more open conformation with a free ADP-ribose-binding cleft (Fig. 2B).

Expression and Purification of the SARS-CoV-2 Macro X domain— An *E. coli* expression vector for the macro X domain was built to express and N-terminally His-tagged protein similar to a SARS-CoV-1 protein studied by Egloff *et al.*[9] Upon induction, a one liter culture of BL21(DE3) cells harboring the vector expresses 50-100 mg of the macro X domain protein that can be purified in one step to apparent homogeneity using immobilized metal affinity chromatography (Fig. 3A).

Nucleotide Binding by the SARS-CoV-2 Macro X domain— Repeated ITC experiments revealed that the purified recombinant protein bound ADP-ribose (Fig. 3C) with a dissociation constant of 10 ± 4 μM (uncertainty is the standard deviation of K_d's from independent titrations). To examine binding specificity, similar titrations were repeated with related nucleotides. The SARS-CoV-2 protein bound ADP, cAMP, ATP and ADP-glucose (Fig. 3D). All nucleotides lacking the ribose moiety bound with similar high affinities, but none bound with an enthalpy change like that seen with ADP-ribose, suggesting specific contacts form between ADP ribose the SARS-CoV-2 protein. Based on what is seen in the SARS-CoV-1 structure (PDB 2fav), these contacts likely occur with the two conserved aspartates (residues 40 and 43 in Fig. 2A). None of the other nucleotides bind with an entropic penalty as seen with ADP-ribose, either, suggesting that the ribose moiety becomes structured when bound to the macrodomain.

The energetics of ADP-ribose binding to the SARS-CoV protein are similar to those seen with the same protein from SARS-CoV-1,[9] MERS-CoV.[8] Enthalpy and entropy of binding were also very similar for all three protein (Fig. 3D). Unlike what is seen with the macro X protein from an alpha coronavirus,[5] enthalpy appears to drive ADP-ribose binding to the macro X domains of the three beta coronaviruses.

To address the molecular basis for ADP-ribose binding, crystallization experiments were performed using the purified SARS-CoV-2 protein and the conditions used to crystallize the homologous protein from SARS-CoV-1.[9] Unfortunately, no crystals formed, nor did crystals form in any of the sparse matrix crystallization screens that were tested. Although both proteins possess the same N-terminal tag, it is possible that it prevents crystallization of the CoV-2 protein, and this idea is presently being tested by subcloning the protein into other expression vectors.

Conclusion— The significance of the study stems mainly from the demonstration that the SARS-CoV-2 macro X domain binds ADP ribose. This is the first step needed to justify screens for potential antivirals that bind in place of ADP ribose. More work needs to be done, however, to understand the antiviral potential of such compounds because the biological role for ADP-ribose binding is still not fully understood. Some work with alpha coronaviruses suggest that ADP-ribose binding by the macro X domain is not needed for viral replication.[10] However, studies with other (+)RNA viruses suggest that macrodomains are essential for virulence.[11] This work is also noteworthy because the synthetic codon-optimized plasmid reported here produces up to 100 mg of soluble macro X domain protein per liter of *E. coli* culture, and this protein retains a high affinity for ADP ribose. The protein could be used for structural studies and screening campaigns. Screening assays with the SARS-CoV-2 protein might actually be more efficient, since the SARS-CoV-2 protein binds ADP-ribose somewhat more tightly (K_d = 10 μM) than the SARS-CoV-1 protein (K_d = 24 μM). The recombinant protein reported here protein might also be useful to others developing SARS-CoV-2 diagnostics.

METHODS

Gene Synthesis— To facilitate comparison between SARS-CoV-1 and SARS-CoV-2, a protein expression vector was generated similar to the one used by Eglott *et al.*[9] To this end, a codon optimized open reading frame was synthesized by GenSript (Piscataway, NJ) that en-

codes the Macro X domain flanked by *Nhe*I and *Bam*H1 restriction sites. This open reading frame was cloned into pET21b to give plasmid pET21-COVID-MacroX. The pET11-COVID-MacroX plasmid was used to transform BL21(DE3) cells.

Protein Purification— Colonies of BL21(DE3) cells harboring the pET21-COVID-MacroX plasmid were used to inoculate 3 ml of LB medium containing 100 mg/ml ampicillin. The starter culture was incubated at 37 °C with shaking at 225 rpm. After the cells grew to an OD_{600} of 1.0, they were transferred to 1 liter of fresh medium containing ampicillin. After the cells reached an OD_{600} of 1.0 again, protein production was induced with 1 mM isopropyl-β-D-thiogalactoside. After growing 16 h at 23 °C, the cells were harvested by centrifugation at 4,000 rpm, 4°C. The resulting cell pellet was suspended in 25 mL of IMAC buffer (20 mM Tris pH 8, 0.5 M NaCl), sonicated on ice for five 1 min bursts, with 2 min rests between, and clarified by centrifugation at 10,000g for 30 min. The supernatant was loaded onto a 5 ml Ni-NTA column and the fractions were eluted with a step gradient from 5 to 500 mM imidazole. Fractions containing the macro X domain protein (5 ml total) were loaded on a 250 ml Sephacryl S300 gel filtration column and eluted with 10 mM MOPS, 150 mM NaCl. Concentration of the purified protein was determined by measuring absorbance at 260 nm using a molar extinction coefficient of 10,555 M^{-1} cm^{-1}, which was calculated with the ProtParam tool (https://web.expasy.org/protparam/).

Isothermal Titration Calorimetry (ITC)— Binding of ADP-ribose to the SARS-CoV-2 macro X domain was measured using a Nano ITC (TA Instruments). Before starting the measurement, samples of both ligand and protein were diluted in 10 mM MOPS, 150 mM NaCl (pH 7) and were degassed at 400 mmHg for 30 minutes. Measurements were taken at 20 °C by injecting 2.0 µl aliquots of 500 µM ADP-Ribose (Sigma) to 50 µM protein (175 µl initial volume) with 250 rpm stirring rate. Using NanoAnalyze Software (v. 3.11.0), data were fitted by non-linear regression to an independent binding model. Briefly, after baseline correction, background heats from ligand-to-buffer titrations were subtracted, and the corrected heats from the binding reaction were used to find best fit parameters for the stoichiometry of the binding (n), free energy of binding (ΔG), apparent enthalpy of binding (ΔH), and entropy change (ΔS). Dissociation constants (K_d) were calculated from the ΔG.

Corresponding Author

*(D.N.F.) Phone: (414) 229-6670. Fax: (414) 229-5530. E-mail: frickd@uwm.edu.

Funding Sources

No competing financial interests have been declared.

ACKNOWLEDGMENTS

The authors are grateful to Matt McCarty, Garrett Breit, Hayden Aristizabal, Trevor R. Melkonian, Dante A. Serrano and Prof. Ionel Popa for valuable technical assistance and helpful discussions.

REFERENCES

1. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. A new coronavirus associated with human respiratory disease in China. Nature 2020;579:265-269.
2. Subissi L, Imbert I, Ferron F, Collet A, Coutard B, Decroly E, Canard B. SARS-CoV ORF1b-encoded nonstructural proteins 12-16: replicative enzymes as antiviral targets. Antiviral Res 2014;101:122-130.
3. 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;484:313-322.
4. Saikatendu KS, Joseph JS, Subramanian V, Clayton T, Griffith M, Moy K, Velasquez J, Neuman BW, Buchmeier MJ, Stevens RC, Kuhn P. Structural basis of severe acute respiratory syndrome coronavirus ADP-ribose-1''-phosphate dephosphorylation by a conserved domain of nsP3. Structure 2005;13:1665-1675.
5. Piotrowski Y, Hansen G, Boomaars-van der Zanden AL, Snijder EJ, Gorbalenya AE, Hilgenfeld R. Crystal structures of the X-domains of a Group-1 and a Group-3 coronavirus reveal that ADP-ribose-binding may not be a conserved property. Protein Sci 2009;18:6-16.
6. Lei J, Kusov Y, Hilgenfeld R. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. Antiviral Res 2018;149:58-74.
7. Leung AKL, McPherson RL, Griffin DE. Macrodomain ADP-ribosylhydrolase and the pathogenesis of infectious diseases. PLoS Pathog 2018;14:e1006864.
8. Cho CC, Lin MH, Chuang CY, Hsu CH. 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;291:4894-4902.
9. Egloff MP, Malet H, Putics A, Heinonen M, Dutartre H, Frangeul A, Gruez A, Campanacci V, Cambillau C, Ziebuhr J, Ahola T, Canard B. Structural and functional basis for ADP-ribose and poly(ADP-ribose) binding by viral macro domains. J Virol 2006;80:8493-8502.
10. Keep S, Bickerton E, Armesto M, Britton P. The ADRP domain from a virulent strain of infectious bronchitis virus is not sufficient to confer a pathogenic phenotype to the attenuated Beaudette strain. J Gen Virol 2018;99:1097-1102.
11. Abraham R, McPherson RL, Dasovich M, Badiee M, Leung AKL, Griffin DE. Both ADP-Ribosyl-Binding and Hydrolase Activities of the Alphavirus nsP3 Macrodomain Affect Neurovirulence in Mice. mBio 2020;11
12. Almeida MS, Johnson MA, Herrmann T, Geralt M, Wüthrich K. Novel beta-barrel fold in the nuclear magnetic resonance structure of the replicase nonstructural protein 1 from the severe acute respiratory syndrome coronavirus. J Virol 2007;81:3151-3161.
13. Johnson MA, Chatterjee A, Neuman BW, Wüthrich K. SARS coronavirus unique domain: three-domain molecular architecture in solution and RNA binding. J Mol Biol 2010;400:724-742.
14. Báez-Santos YM, Barraza SJ, Wilson MW, Agius MP, Mielech AM, Davis NM, Baker SC, Larsen SD, Mesecar AD. X-ray structural and biological evaluation of a series of potent and highly selective inhibitors of human coronavirus papain-like proteases. J Med Chem 2014;57:2393-2412.

15. Serrano P, Johnson MA, Chatterjee A, Neuman BW, Joseph JS, Buchmeier MJ, Kuhn P, Wüthrich K. Nuclear magnetic resonance structure of the nucleic acid-binding domain of severe acute respiratory syndrome coronavirus nonstructural protein 3. J Virol 2009;83:12998-13008.

16. Akaji K, Konno H, Mitsui H, Teruya K, Shimamoto Y, Hattori Y, Ozaki T, Kusunoki M, Sanjoh A. Structure-based design, synthesis, and evaluation of peptide-mimetic SARS 3CL protease inhibitors. J Med Chem 2011;54:7962-7973.

17. Kirchdoerfer RN, Ward AB. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. Nat Commun 2019;10:2342.

18. Egloff MP, Ferron F, Campanacci V, Longhi S, Rancurel C, Dutartre H, Snijder EJ, Gorbalenya AE, Cambillau C, Canard B. The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. Proc Natl Acad Sci U S A 2004;101:3792-3796.

19. Ma Y, Wu L, Shaw N, Gao Y, Wang J, Sun Y, Lou Z, Yan L, Zhang R, Rao Z. Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. Proc Natl Acad Sci U S A 2015;112:9436-9441.

20. Jia Z, Yan L, Ren Z, Wu L, Wang J, Guo J, Zheng L, Ming Z, Zhang L, Lou Z, Rao Z. Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis. Nucleic Acids Res 2019;47:6538-6550.

21. Ricagno S, Egloff MP, Ulferts R, Coutard B, Nurizzo D, Campanacci V, Cambillau C, Ziebuhr J, Canard B. Crystal structure and mechanistic determinants of SARS coronavirus nonstructural protein 15 define an endoribonuclease family. Proc Natl Acad Sci U S A 2006;103:11892-11897.

22. Decroly E, Debarnot C, Ferron F, Bouvet M, Coutard B, Imbert I, Gluais L, Papageorgiou N, Sharff A, Bricogne G, Ortiz-Lombardia M, Lescar J, Canard B. Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. PLoS Pathog 2011;7:e1002059.

23. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605-1612.

24. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Protein Sci 2016;86:2.9.1-2.9.37.

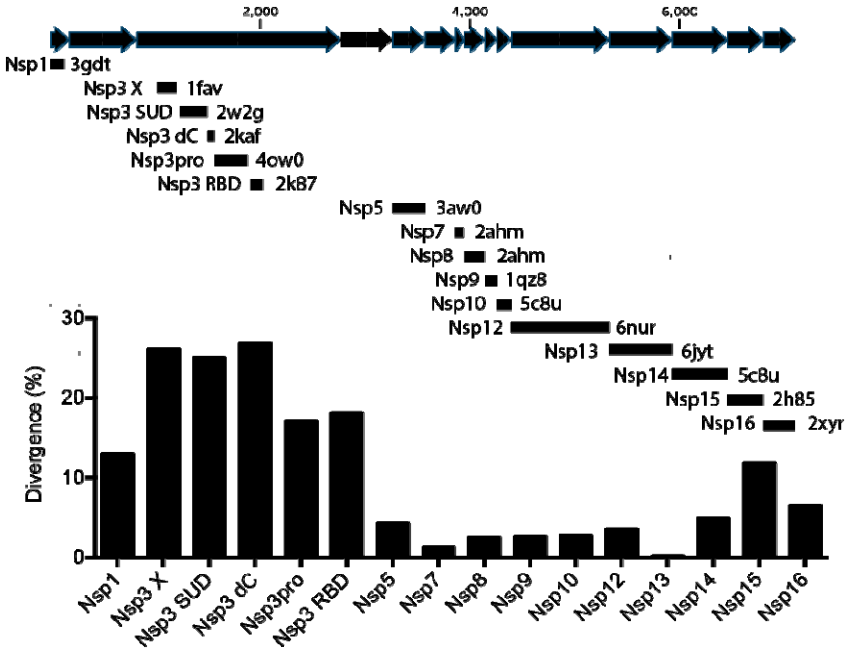


FIGURE 1. **Sequence divergence between potential drug targets in SARS-CoV-1 and SARS-CoV-2.** The SARS-CoV-2 rep 1ab peptide sequence was aligned with each of the PDB files listed, which describe an atomic structure of a homologous region of the SARS-CoV-1 rep 1ab polyprotein. Nsp's are shown in sequence as black arrows (note: there is no "nsp11", and the translational frameshift occurs after nsp10). The percent of amino that differ in each protein is plotted. For details see Fig. S1 (supporting information). Nsp1 is and interferon antagonist.[12] The Nsp3 X domain is studied here, the Nsp3 SUD consists of tandem macrodomains that bind G-quadruplex structures,[3] Nsp dC is the C-terminus of the SUD,[13] Nsp3pro is a papain-like protease,[14] and Nsp3 RBD is another possible RNA binding domain.[15] Nsp5 is the main viral protease.[16] Nsp7 and Nsp8 are polymerase cofactors.[17] Nsp9 is an RNA binding protein.[18] Nsp10 is a zinc-binding cofactor for Nsp14 and Nsp16.[19] Nsp12 is the RNA polymerase.[17] Nsp13 is a helicase.[20] Nsp14 is a 3'-5' exonuclease and a 7-methyltransferase.[19] Nsp15 is an RNA endonuclease.[21] Nsp16 is an RNA cap 2'-O-Methyltransferase.[22]

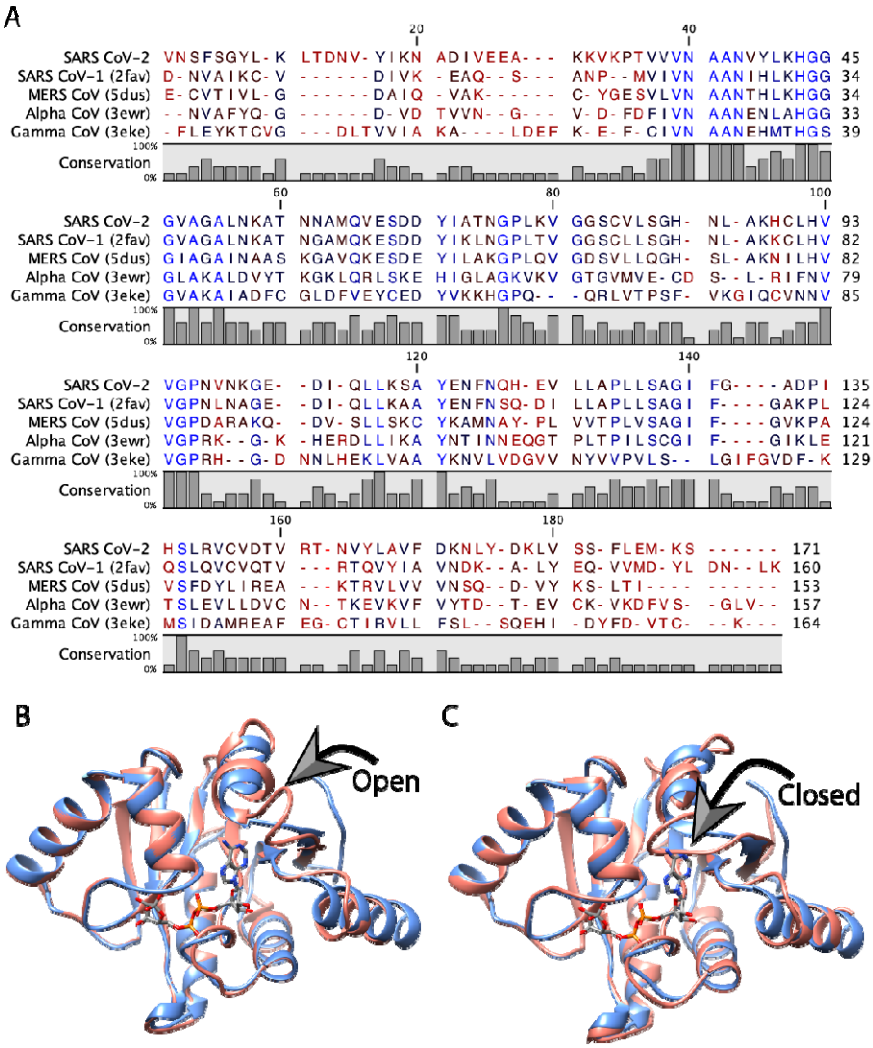


FIGURE 2. **Variation in the macro X domains of coronaviruses.** (A) Macro X domain structures were aligned using the "MatchMaker" function of UCSF Chimera (v. 1.14)[23] along with a hypothetical structure of the SARS CoV-2 macro X domain generated using and the MODELLER[24] plugin for UCSF Chimera. PDB file 2av was used to guide modeling. Alignment is shown above a conservation graph generated by CLC Sequence Viewer (7.7.1) with amino acid colored in a gradient from blue (identical) to red (poorly conserved). (B) SARS-CoV-1 macro X domain (PDB file 2fav, blue) aligned with modeled SARS-CoV-2 macro X domain (red), in which the variable N-terminus folds so that ADP-ribose might bind (zDOPE score -0.44). (B) Model of the SARS-CoV-2 macro domain in which the N-terminus blocks ADP-ribose binding (zDOPE score -0.61). In (B) and (C) the ADP-ribose bound to SARS-CoV-1 is shown as sticks.

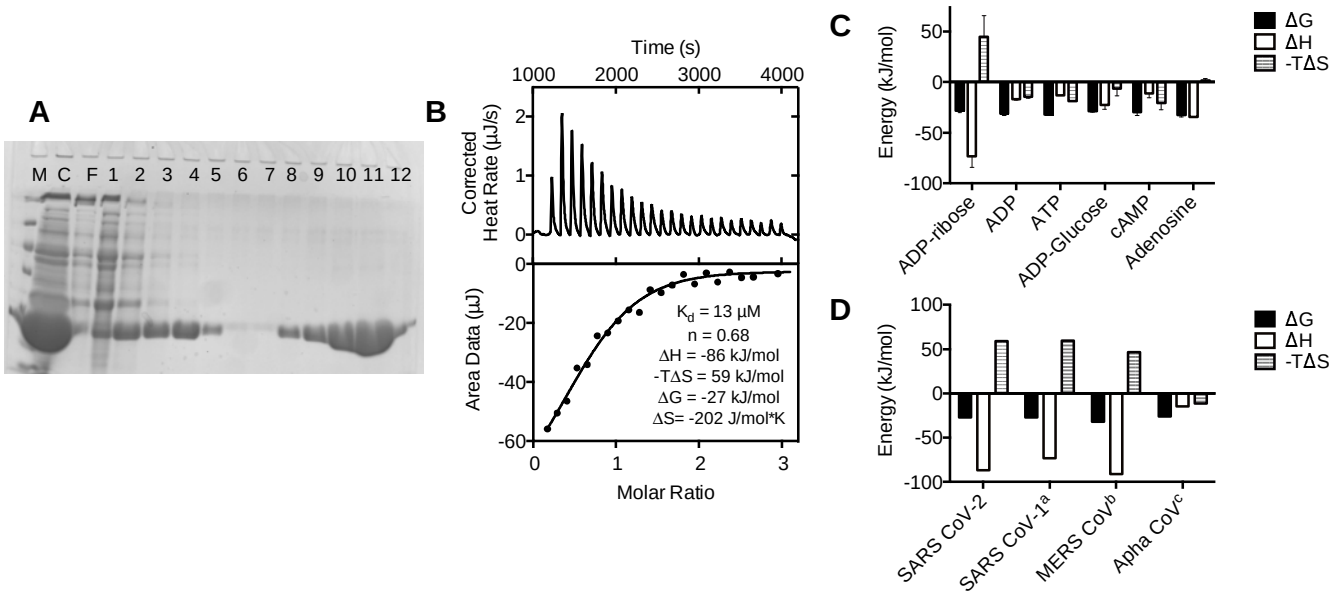
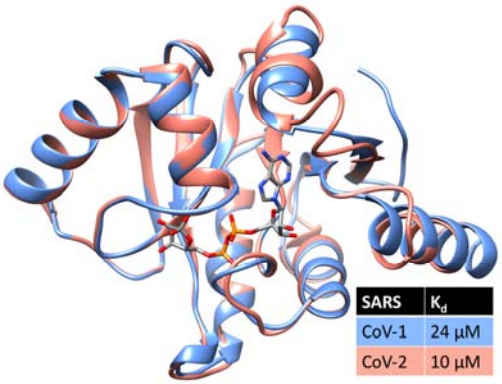


FIGURE 3. **The SARS-CoV-2 macro X domain binds ADP ribose.** (A) 15% SDS PAGE showing 10 µL samples of: a soluble crude lysate of induced BL21(DE3) cells harboring the plasmid p21-COVID-MacroX (lane C), proteins the do not bind a Ni-NTA column (F), and fractions eluted from a NiNTA column during an imidazole step gradient from 0 mM (lanes 1-3), 5 mM (lanes 4-6), 40 mM (lanes 7-9), and 500 mM (lanes 10-12). Protein markers (lane M) are 116, 66.2, 45, 35, and 25 kDa. (B) Example ITC experiment in which purified SARS-CoV-2 was titrated with ADP ribose. (C) ITC experiments like those shown in panel B were repeated tree times with each of the compounds listed. Means are plotted and error bars are standard deviations. Average (±SD) dissociation constants were 10 ± 4 µM for ADP-ribose, 8 ± 9 µM for ADP, 3 ± 3 µM for ATP, 6± 4 µM for ADP-glucose, 2 ± 1 µM for cAMP, and 2 ± 1 µM for adenosine. (D) Comparison of the thermodynamics of ADP-ribose binding by macro X domains from SARS-CoV-2 (data from panel C), SARS-CoV-1, MERS-CoV, and an alpha coronavirus. [a]Data from Egloff et al.[18] [b]Data Cho et al.[8] [c]Data from Piotrowski et al.[5]


SARS
K_d
CoV-1
24 μM
CoV-2
10 μM