

## Fragment Binding to the Nsp3 Macrodome of SARS-CoV-2 Identified Through Crystallographic Screening and Computational Docking

Marion Schuller<sup>1†</sup>, Galen J. Correy<sup>2†</sup>, Stefan Gahbauer<sup>3†</sup>, Daren Fearon<sup>4†</sup>, Taiasean Wu<sup>5,6</sup>, Roberto Efraín Díaz<sup>2,7</sup>, Iris D. Young<sup>2,8</sup>, Luan Carvalho Martins<sup>9</sup>, Dominique H. Smith<sup>10</sup>, Ursula Schulze-Gahmen<sup>8</sup>, Tristan W. Owens<sup>8</sup>, Ishan Deshpande<sup>8</sup>, Gregory E. Merz<sup>8</sup>, Aye C. Thwin<sup>8</sup>, Justin T. Biel<sup>8</sup>, Jessica K. Peters<sup>8</sup>, Michelle Moritz<sup>8</sup>, Nadia Herrera<sup>8</sup>, Huong T. Kratochvil<sup>8</sup>, QCRG Structural Biology Consortium<sup>8‡</sup>, Anthony Aimon<sup>4</sup>, James M. Bennett<sup>11</sup>, Jose Brandao Neto<sup>4</sup>, Aina E. Cohen<sup>12</sup>, Alexandre Dias<sup>4</sup>, Alice Douangamath<sup>4</sup>, Louise Dunnett<sup>4</sup>, Oleg Fedorov<sup>11</sup>, Matteo P. Ferla<sup>13</sup>, Martin Fuchs<sup>14</sup>, Tyler J. Gorrie-Stone<sup>4</sup>, James M. Holton<sup>15,16,12</sup>, Michael G. Johnson<sup>17</sup>, Tobias Krojer<sup>11,18</sup>, George Meigs<sup>15,16</sup>, Ailsa J. Powell<sup>4</sup>, Johannes Gregor Matthias Rack<sup>1</sup>, Victor L Rangel<sup>11,18,19</sup>, Silvia Russi<sup>12</sup>, Rachael E. Skyner<sup>4</sup>, Clyde A. Smith<sup>12</sup>, Alexei S. Soares<sup>20</sup>, Jennifer L. Wierman<sup>12</sup>, Kang Zhu<sup>1</sup>, Natalia Jura<sup>21</sup>, Alan Ashworth<sup>10</sup>, John Irwin<sup>3</sup>, Michael C. Thompson<sup>22</sup>, Jason E. Gestwicki<sup>3,5</sup>, Frank von Delft<sup>4,11,18,23\*</sup>, Brian K. Shoichet<sup>3\*</sup>, James S. Fraser<sup>2\*</sup>, Ivan Ahel<sup>1\*</sup>

### Affiliations

- <sup>1</sup> Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK
- <sup>2</sup> Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, CA, USA
- <sup>3</sup> Department of Pharmaceutical Chemistry, University of California San Francisco San Francisco, CA, USA
- <sup>4</sup> Diamond Light Source Ltd., Harwell Science and Innovation Campus, Didcot OX11 0DE, United Kingdom
- <sup>5</sup> Institute for Neurodegenerative Disease, University of California San Francisco, CA, USA
- <sup>6</sup> Chemistry and Chemical Biology Graduate Program, University of California San Francisco, CA, USA
- <sup>7</sup> Tetrad Graduate Program, University of California San Francisco, CA, USA
- <sup>8</sup> Quantitative Biosciences Institute (QBI) Coronavirus Research Group Structural Biology Consortium, University of California San Francisco, CA, USA
- <sup>9</sup> Biochemistry Department, Institute for Biological Sciences, Federal University of Minas Gerais. Belo Horizonte, Brazil
- <sup>10</sup> Helen Diller Family Comprehensive Cancer, University of California San Francisco, CA, USA
- <sup>11</sup> Centre for Medicines Discovery, University of Oxford, South Parks Road, Headington, OX3 7DQ, UK
- <sup>12</sup> Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Center, Menlo Park, CA 94025, USA
- <sup>13</sup> Wellcome Centre for Human Genetics, University of Oxford, Old Road Campus, Oxford OX3 7BN, UK
- <sup>14</sup> National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, NY, USA
- <sup>15</sup> Department of Biochemistry and Biophysics, University of California San Francisco, CA, USA
- <sup>16</sup> Department of Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
- <sup>17</sup> ChemPartner Corporation, South San Francisco, CA, USA
- <sup>18</sup> Structural Genomics Consortium, University of Oxford, Old Road Campus, Roosevelt Drive, Headington OX3 7DQ, UK
- <sup>19</sup> School of Pharmaceutical Sciences of Ribeirao Preto, University of Sao Paulo, São Paulo, Brazil
- <sup>20</sup> Photon Sciences, Brookhaven National Laboratory, Upton, NY, USA
- <sup>21</sup> Department of Cellular and Molecular Pharmacology, University of California San Francisco, CA, USA
- <sup>22</sup> Department of Chemistry and Chemical Biology, University of California Merced, CA, USA
- <sup>23</sup> Department of Biochemistry, University of Johannesburg, Auckland Park, 2006, South Africa

† These authors contributed equally

‡ Full listing at the end of the manuscript

\* To whom correspondence should be addressed; Email: [ivan.ahel@path.ox.ac.uk](mailto:ivan.ahel@path.ox.ac.uk), [jfraser@fraserlab.com](mailto:jfraser@fraserlab.com), [bshoichet@gmail.com](mailto:bshoichet@gmail.com), [frank.von-delft@diamond.ac.uk](mailto:frank.von-delft@diamond.ac.uk)

## SUPPLEMENTARY FIGURES

**Supplementary Figure 1.** Ultra-high resolution features in Mac1 electron density maps.

**Supplementary Figure 2.** Comparison of isomorphism and DMSO tolerance of the C2 and P4<sub>3</sub> crystals.

**Supplementary Figure 3.** Crystal packing determines active site accessibility.

**Supplementary Figure 4.** Structure and sequence comparison of Mac1 with related viral and human macrodomains.

**Supplementary Figure 5.** Physical properties, scaffold and chemotype analysis of screened fragment libraries.

**Supplementary Figure 6.** Overview of fragment binding to protomer A and protomer B of the P4<sub>3</sub> crystals.

**Supplementary Figure 7.** Additional soaking hits from docking and adenine-N3 vs -N9-alkylated isomers.

**Supplementary Figure 8.** Mac1 subsites compared to the adenine binding site in kinases and the oxyanion binding site in carboxylesterases.

**Supplementary Figure 9.** Comparison of DSF, HTRF, and ITC results for compounds tested in all assays.

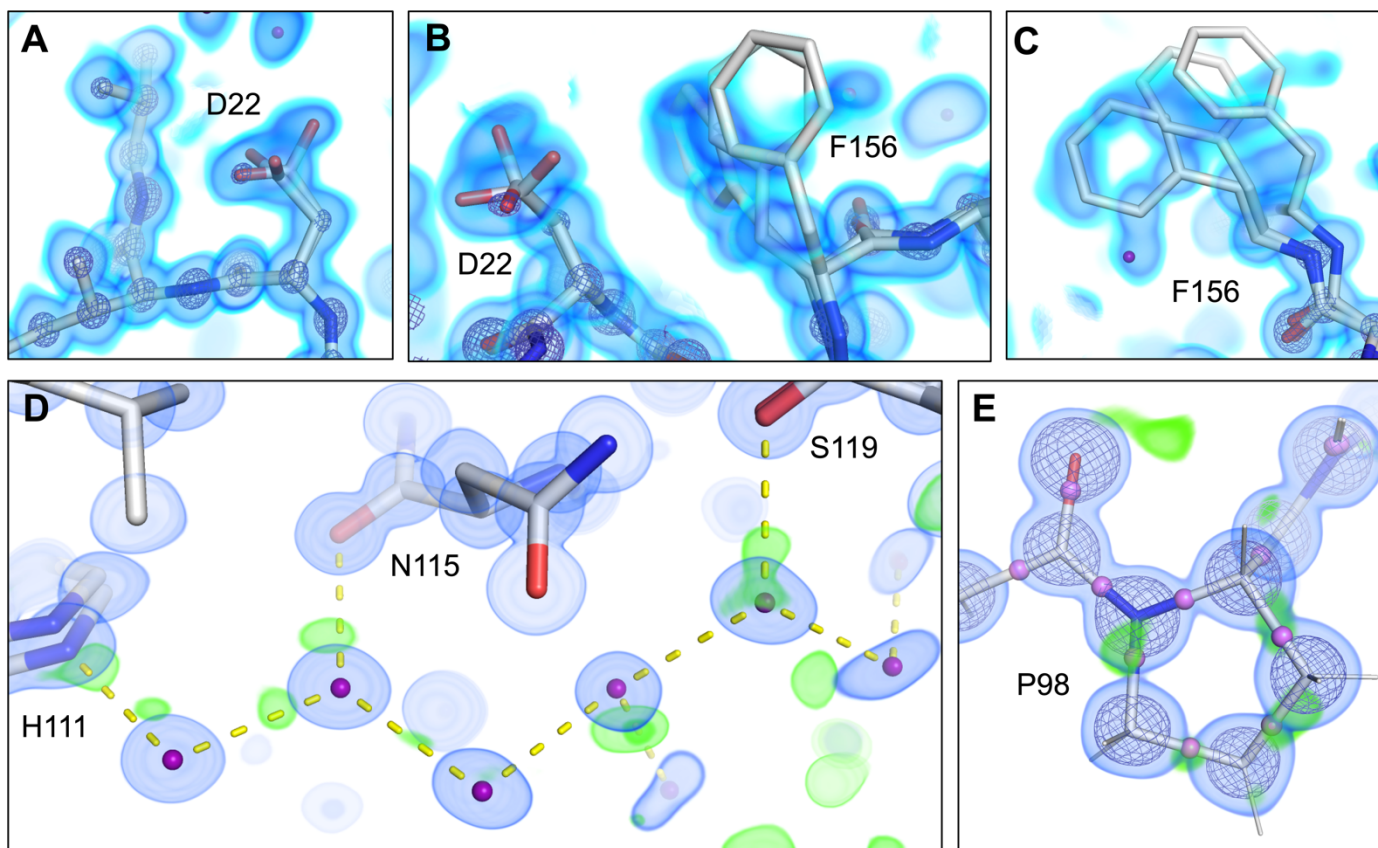
## SUPPLEMENTARY TABLE

**Supplementary Table 1.** Characterisation of docking and crystallographic hits including docking scores and solution binding data (spreadsheet, **SI\_table1**)

## SUPPLEMENTARY DATA SETS

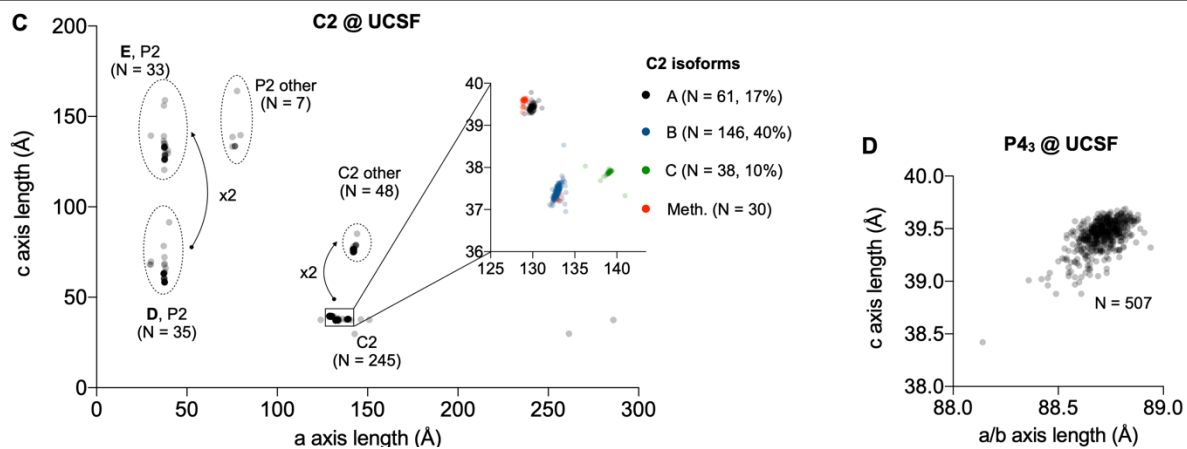
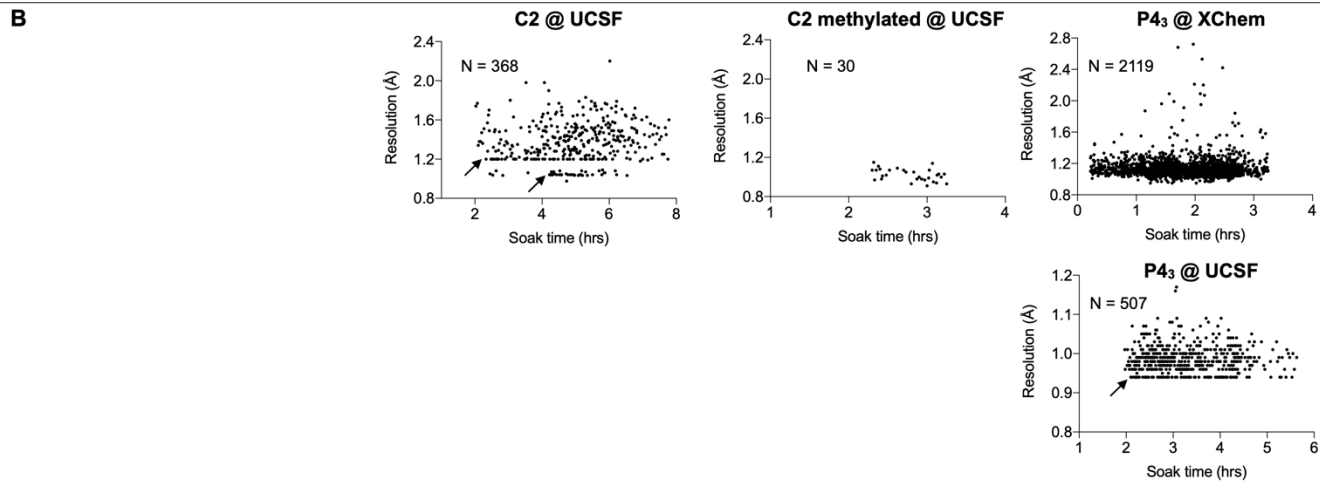
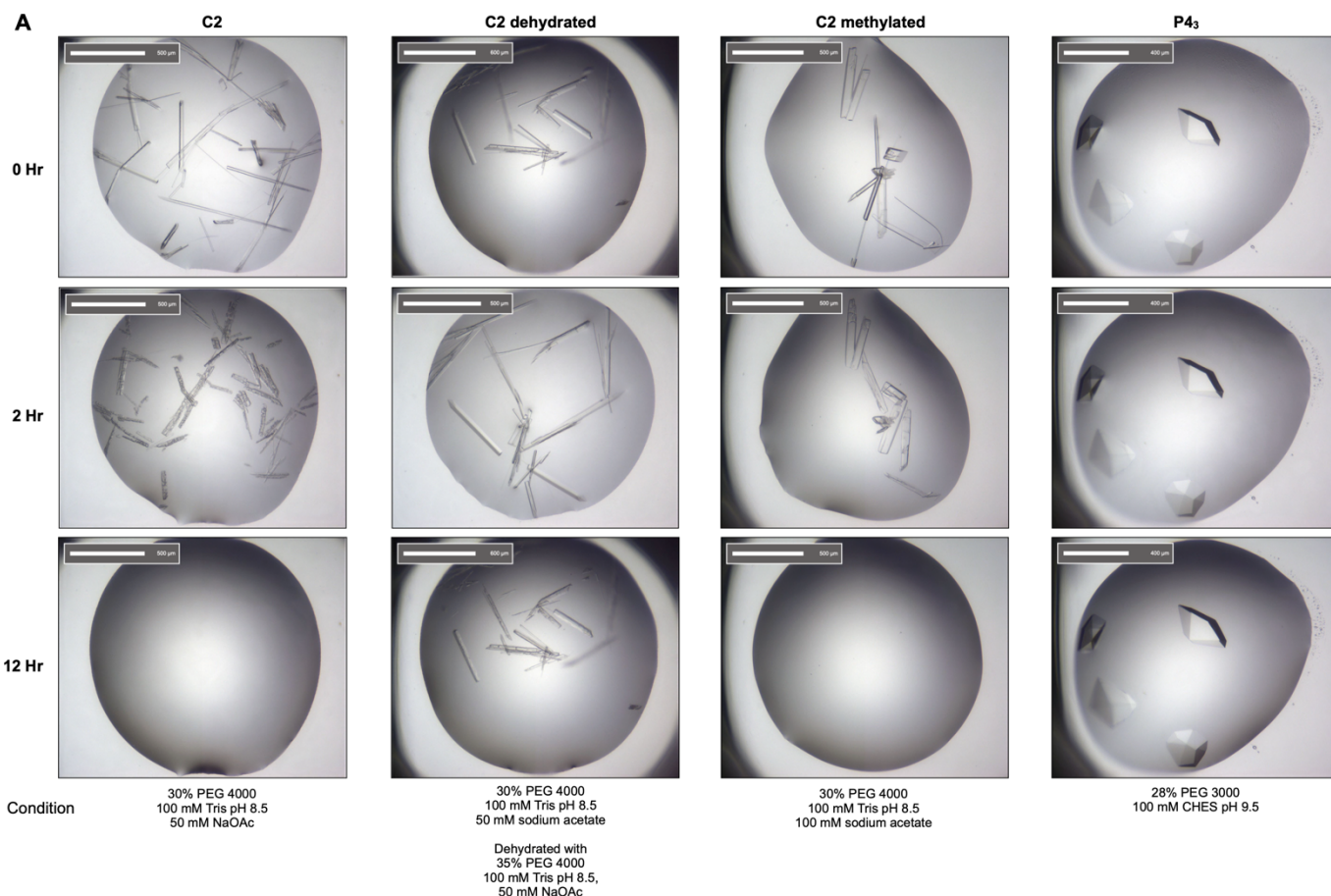
**Supplementary data set 1.** Crystallographic fragment screening information including sequences, data collection strategy, data reduction and refinement stats, summary of all fragment datasets, subsite classification (spreadsheet, **SI\_dataset\_1**)

**Supplementary data set 2.** Selected fragment structures. DSF, ITC, and peptide displacement data for all tested compounds (PDF, **SI\_dataset\_2**).

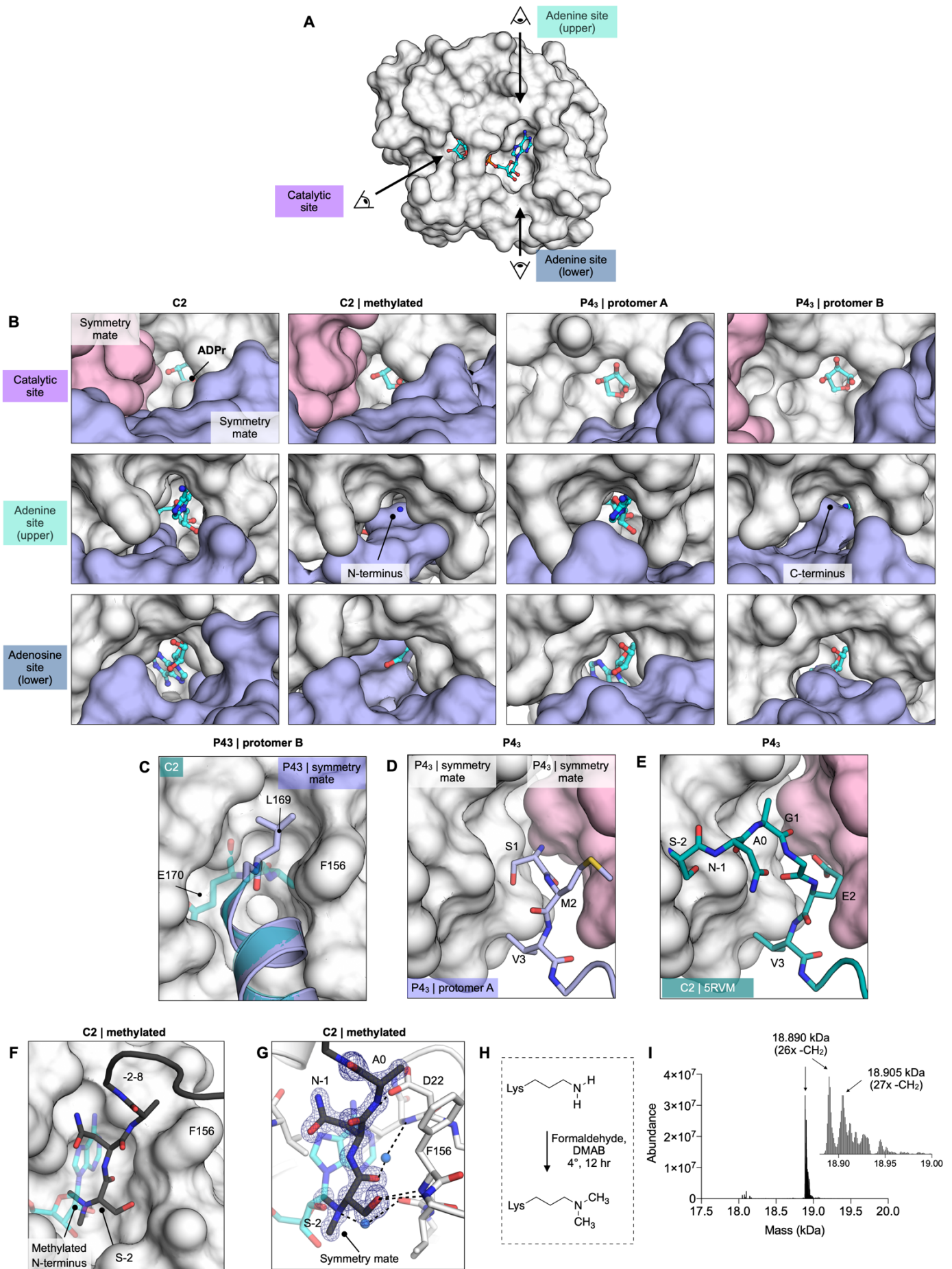


**Figure S1. Ultra-high resolution features in Mac1 electron density maps.**

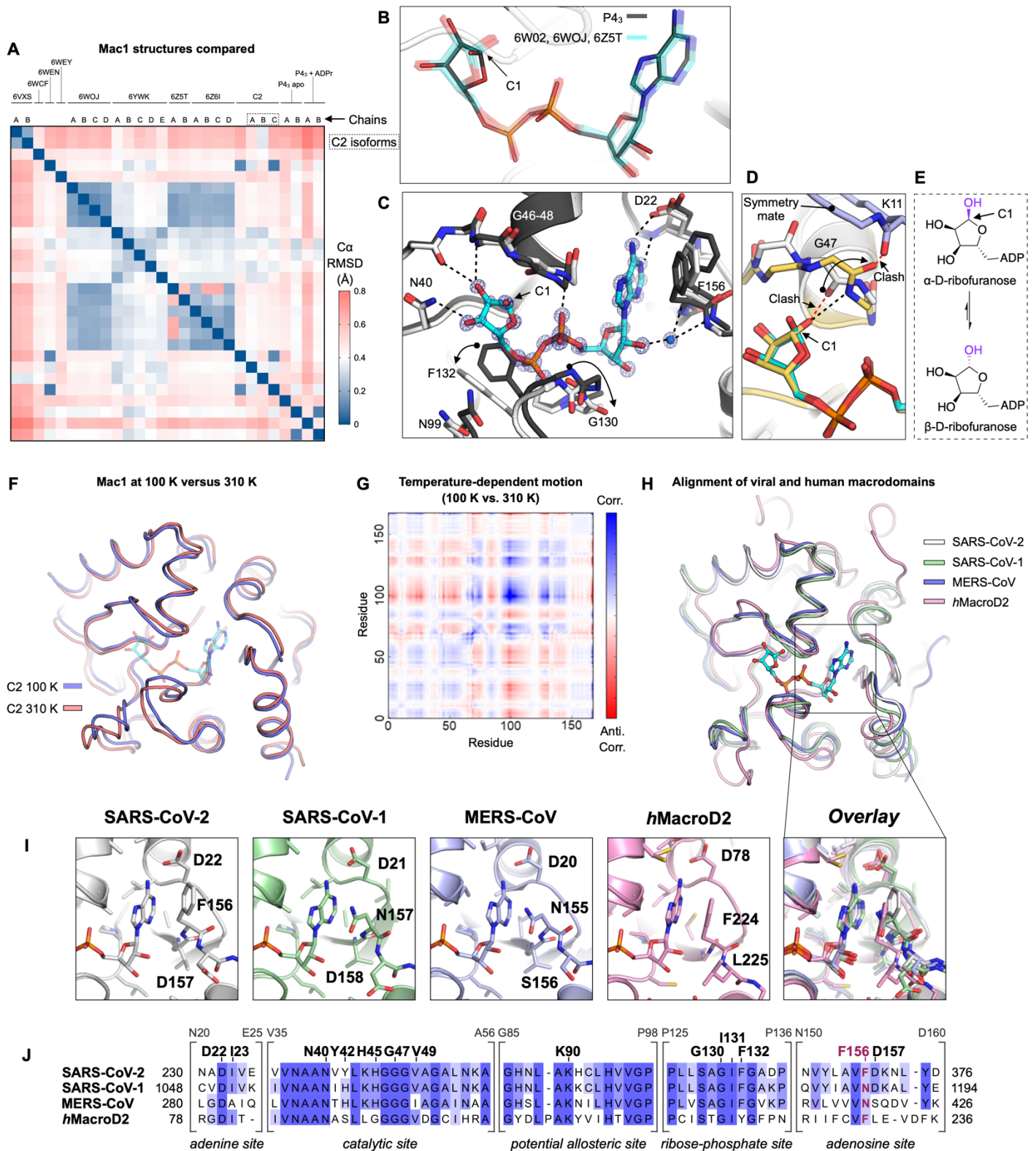
**A,B,C)** Residues Asp22 and Phe156 both display conformational heterogeneity in high-resolution electron density maps of apo Mac1. The ultra-high resolution nature of the electron density is evident in (A), where atoms belonging to residues Ile23 and Leu24, as well as the backbone, appear as separate spheres at high contour. In (A,B,C) three  $2mF_o - Df_c$  maps are shown, contoured at  $4.0 \sigma$  (blue mesh),  $1.0 \sigma$  (blue volume), and  $0.5 \sigma$  (cyan volume). **D)** Mac1 electron density reveals ordered water networks, including difference density that suggests the positions of hydrogen atoms within the water network. In (D), a  $2mF_o - Df_c$  map is shown, contoured at  $1.5 \sigma$  (blue volume), and a  $mF_o - Df_c$  map is also shown, contoured at  $2.5 \sigma$  (green volume). Hydrogen bonding interactions are depicted as dashed yellow lines. **E)** Two electron density maps are shown for P98, a  $2mF_o - Df_c$  map, contoured at  $1.5 \sigma$  (blue volume), and a  $mF_o - Df_c$  map, contoured at  $2.0 \sigma$  (green volume). The green peaks of positive difference density between heavy atom positions suggest covalent bonding densities, and we refined a model of Mac1 that included interatomic scatterers (IAS), shown in violet in the figure.



**Figure S2. Comparison of isomorphism and DMSO tolerance of the C2 and P4<sub>3</sub> crystals.** **A)** Images of crystals after soaking with 10% DMSO for 0, 2 and 12 hours. **B)** Resolution of the three crystal forms as a function of soak time for the datasets collected at XChem and UCSF. The arrows indicate where the measurement of high resolution reflections was limited by the experimental setup. **C)** Multiple isoforms were observed for the C2 crystals after dehydration. Isoforms were distinguished based on differences in the a and c unit cell lengths. Arrows indicate where doubling of the a or c axis occurred. Inset: the majority of the datasets that were indexed in C2 (245, 84%) could be clustered into three isoforms (A, B and C). Of the 30 datasets collected for crystals grown from methylated protein, the majority (28) were similar to the A isoform. **D)** The P4<sub>3</sub> crystals were isomorphous.



**Figure S3. Crystal packing in Mac1 crystals determines active site accessibility.** **A)** Active site access in the C2 and P4<sub>3</sub> crystals. Mac1 is shown as a white surface with ADP-ribose bound in the active site shown as cyan sticks. The three access points are indicated with arrows. **B)** Crystal packing defines the three access points. The catalytic site is partially obstructed in the C2 crystals, but open in both protomers of the P4<sub>3</sub> crystals. In both the methylated C2 crystals and protomer B of the P4<sub>3</sub> crystals, the adenosine site is obstructed. **C)** The C-terminal leucine (Leu169) of the P4<sub>3</sub> construct occupies the adenosine site of a symmetry mate. The adenosine site is shown as a white surface and the C-terminal residues with blue sticks/cartoon. The C2 sequence (transparent teal cartoon/sticks) has an additional residue at the C-terminus (Glu170), and is therefore incompatible with the P4<sub>3</sub> crystal packing. **C)** The N-terminal residues of the P4<sub>3</sub> sequence (blue sticks) pack between two symmetry mates (white and pink surface). Compared to the P4<sub>3</sub> sequence, the C2 sequence contains a substitution (Met2Glu) and a three residue insertion (Asn-Ala-Gly). These residues were typically disordered, however, they were resolved in one of the fragment structures (ZINC157088 | 5RVM) (shown aligned to the P4<sub>3</sub> protomer A in **E)**). Like differences in the C-termini, differences in the N-termini may have contributed to the distinct crystal packing seen for the two Mac1 structures reported in this work. **F)** The adenosine site was obstructed by a symmetry mate in the structure determined from crystals grown using methylated C2 protein. **G)** In the structure of methylated Mac1, the side-chain hydroxyl of Ser-2 occupies the oxyanion subsite. Electron density (2mF<sub>o</sub>-DF<sub>c</sub>) is shown as a blue mesh, contoured at 1.5  $\sigma$ . **H)** Free amines were methylated using formaldehyde and dimethylamine borane (DMAB). The reaction is shown for lysine, however, based on the electron density shown in **G)**, the N-terminal amine was methylated as well. The methylated amines would be protonated at the pH used to grow crystals (pH 8.5). **I)** LC/MS analysis of methylated Mac1 (C2 construct). The mass spectrum was deconvoluted using MaxEnt1. The major peak (18.89 kDa) is consistent with the methylation of 13 lysine residues (26x -CH<sub>2</sub>). The minor peak (18.905 kDa + 15 Da) suggests that methylation was not 100% complete.

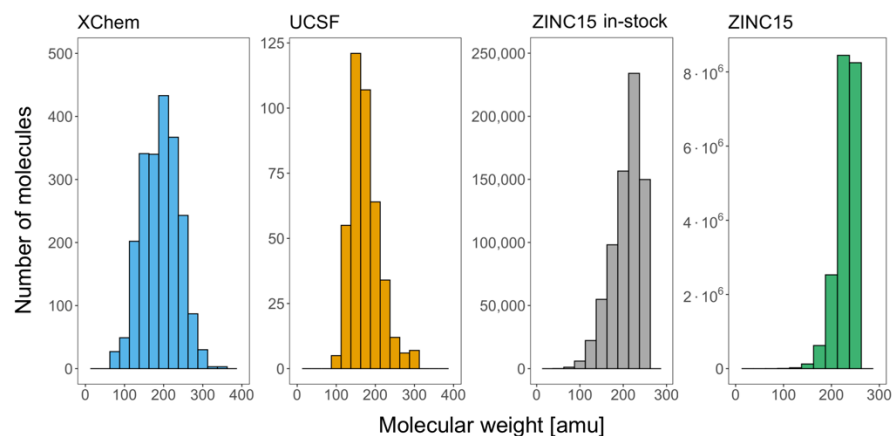


**Figure S4. Structure and sequence comparison of Mac1 with related viral and human macrodomains. A)** Heatmap showing the C $\alpha$  RMSD values after C $\alpha$  alignment for 10 previously reported SARS-CoV-2 Mac1 structures (6VXS, 6W02, 6W6Y, 6WCF, 6WEN, 6WOJ, 6WEY, 6YWK, 6Z5T, 6Z6I) and the new structures reported in this work. **B)** Stick representation showing the previously reported Mac1-ADPr structures (cyan sticks) and the new structure determined using P<sub>43</sub> crystals (grey sticks). The agreement between ADP-ribose

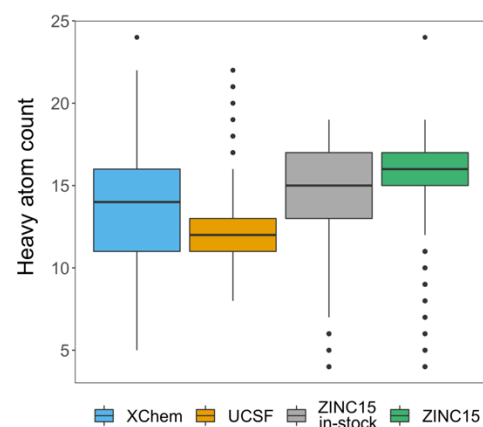


is excellent, despite different anomers of the terminal ribose being present ( $\alpha$  in the  $P4_3$  structure,  $\beta$  in the previously reported structures). **C)** The structural changes previously reported to occur upon ADPr binding are captured by the Mac1-ADPr structure determined using  $P4_3$  crystals. The apo  $P4_3$  structure is shown with dark gray sticks, with arrows indicating the changes in protein conformation upon ADP-ribose binding (white sticks). Electron density ( $2mF_o-DF_c$ ) is contoured at  $4\sigma$  (blue mesh). **D)** The  $\alpha$ -anomer of the terminal ribose of ADP-ribose was observed in the  $P4_3$  crystal form (cyan and white sticks). In previously reported structures (e.g. PDB 6W02, yellow sticks), a flip in Gly47 allows the  $\beta$ -anomer to bind by removing a steric block (red dashed line) and forming a new H-bond (black dashed line). However, the Gly47 flip is incompatible with the  $P4_3$  crystal form because it would clash with the Lys11 carbonyl of a symmetry mate (blue sticks). In  $\alpha$ -anomer, the anomeric hydroxyl is orientated away from Gly47, and binding can proceed without the peptide flip. **E)** Interconversion between ribose anomers in solution. **F)** Comparison of SARS-CoV-2 Nsp3 Mac1 structures at 100 K (blue) and 310 K (red). The adenosine diphosphoribose ligand shown in the figure (cyan) is modeled according to its position in PDB 6W02. **G)** Correlation plot showing structural differences between the 100 K and 310 K structures. To generate the plot, the 100 K and 310 K structures were aligned and difference vectors were calculated between identical C $\alpha$  atoms in the two structures. The plot shows all pairwise dot products between these difference vectors, revealing the extent to which temperature-dependent changes are correlated across the structure. Positive dot products (positive correlations) are colored blue, while dot products (negative correlations) are shown in red. The pattern of positive and negative correlations is characteristic of a hinge-bending motion. **H)** Alignment of three coronavirus macrodomain structures with a human macrodomain (hMacroD2). ADPr from the SARS-CoV-2 structure is shown with cyan sticks. **I)** Comparison of the adenosine binding site highlighting key residues involved in adenine and fragment interaction. The adenine coordination by Phe156 is unique to SARS-CoV-2 amongst betacoronaviruses and replaced in SARS-CoV-1 (PDB 2FAV) and MERS-CoV (PDB 5HOL) with asparagine. Human macrodomains including MacroD2 (PDB 4IQY) interact with adenine as SARS-CoV-2 Nsp3 Mac1 with a phenylalanine in this position which needs to be considered for achieving inhibitor selectivity for viral over human macrodomains. **J)** Sequence alignment showing conservation of residues in the ADP-ribose, catalytic and potential allosteric sites which are targeted by the fragments. Residue numbers on top refer to the construct residue numbering of SARS-CoV-2 Nsp3 Mac1. Numbers on either end of the alignment are residue numbers in the full-length proteins. The adenine-coordinating Phe156 is highlighted in red.

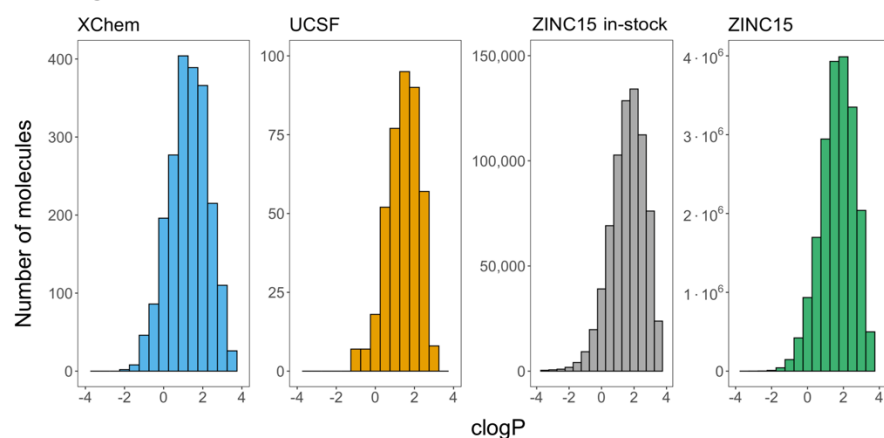
### A Molecular weight



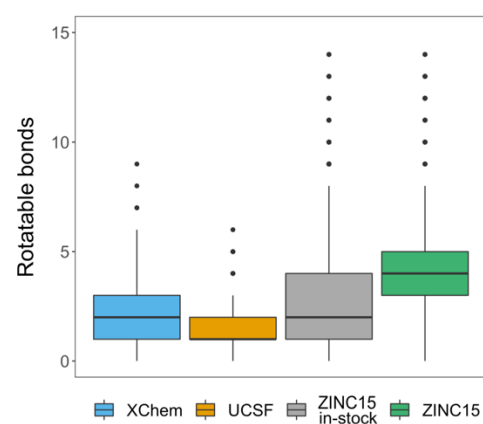
### C Number of heavy atoms



### B cLogP



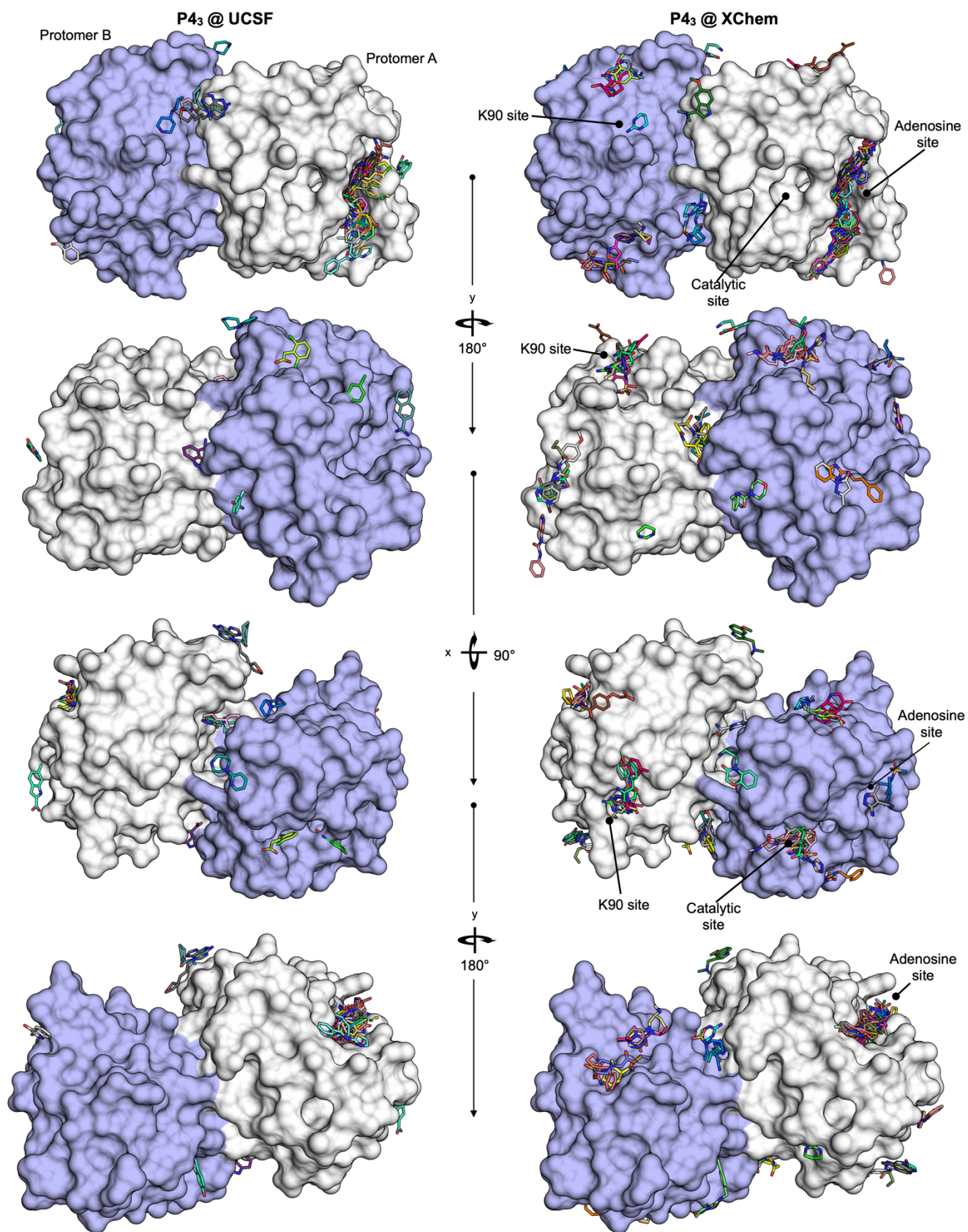
### D Number of rotatable bonds



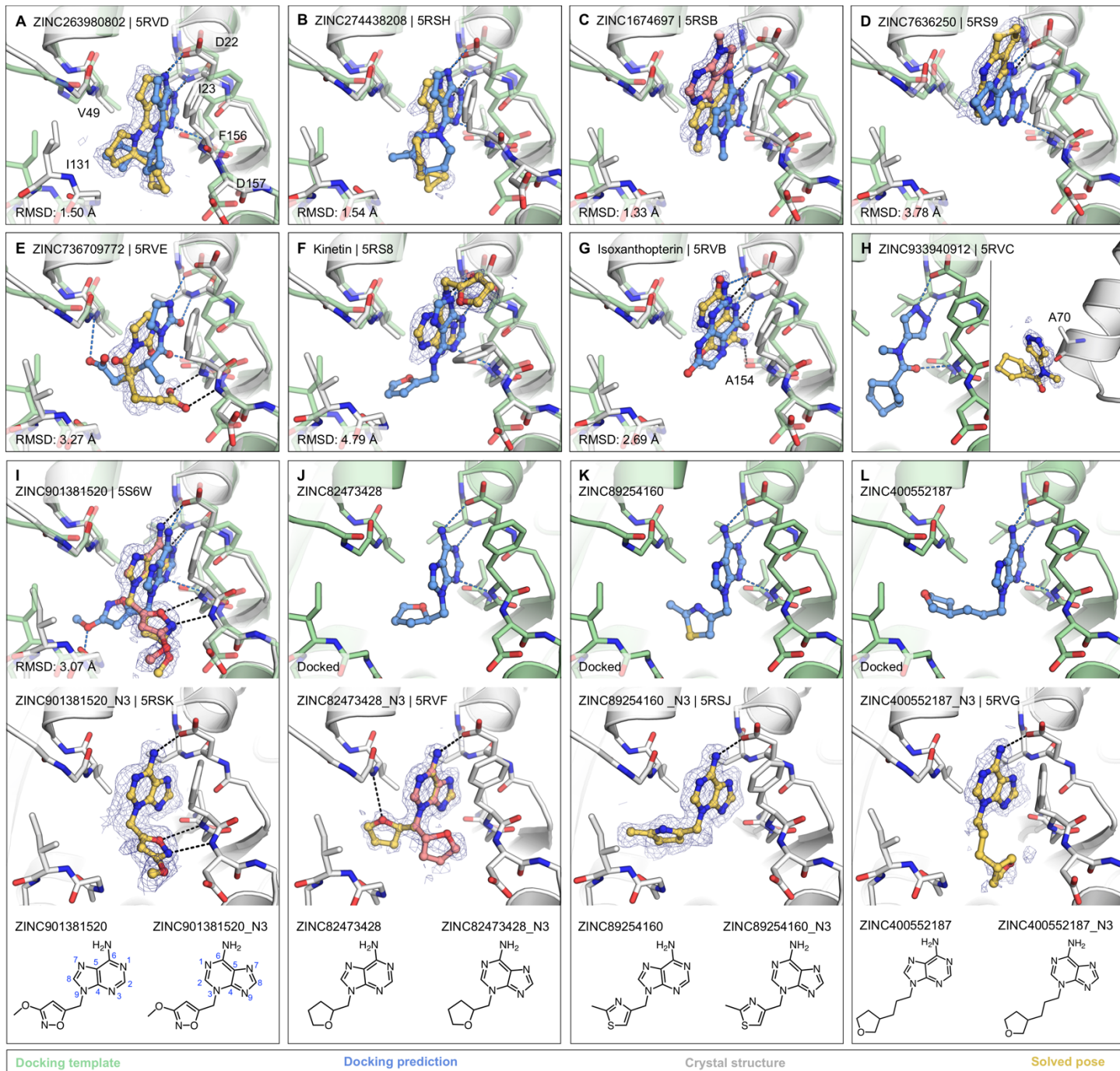
### E Scaffold and chemotype analysis

Fragment Library	Number of molecules	Distinct Bemis-Murcko scaffolds	Number of pyrimidines	Number of anions
XChem	2,126	809	72 (3.39%)	115 (5.4%)
UCSF	411	179	12 (2.92%)	144 (35%)
ZINC15 (in-stock)	722,963	69,244	41,531 (5.74%)	85,398 (11.8%)
ZINC15	20,006,175	803,333	890,199 (4.44%)	739,184 (3.7%)

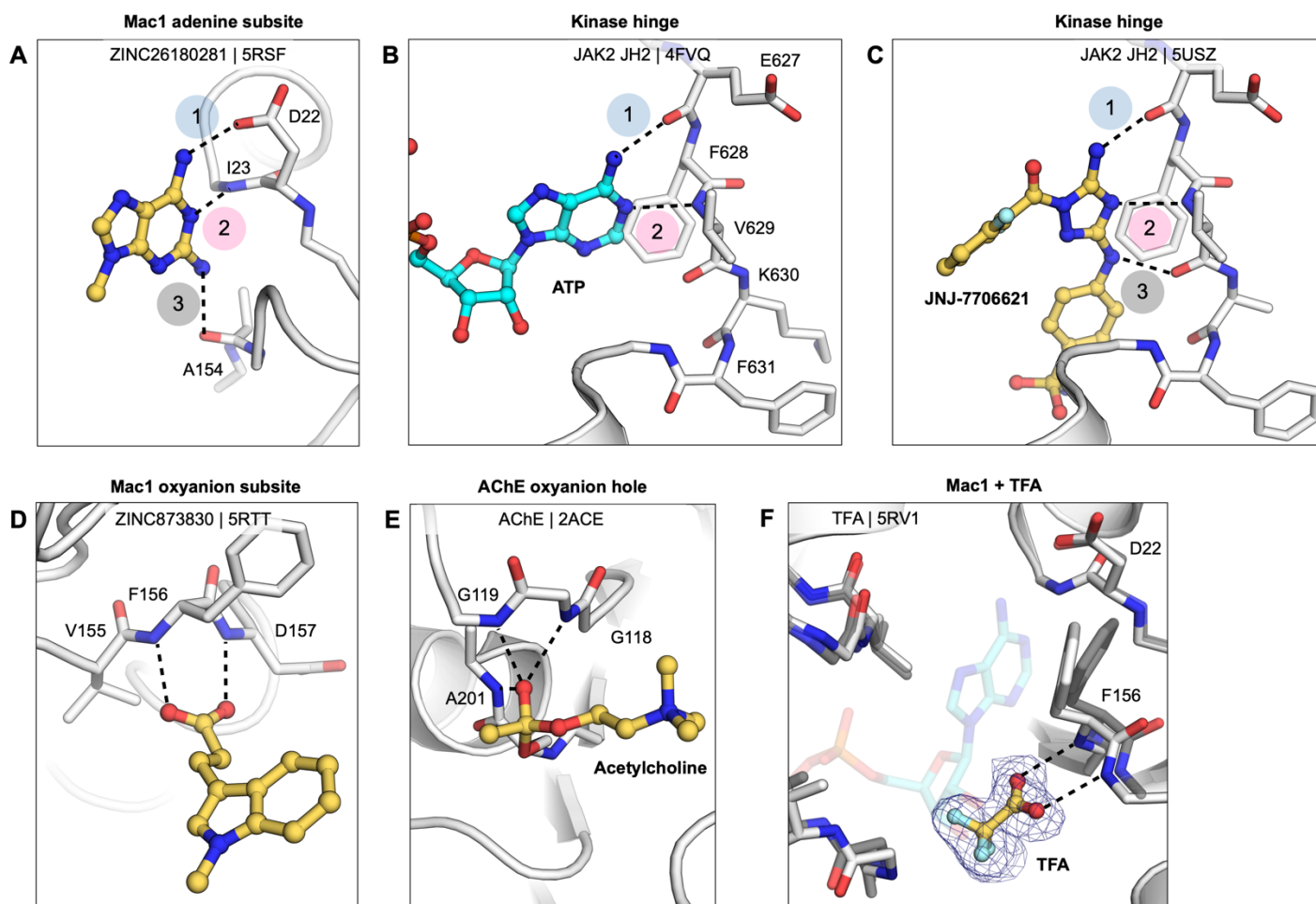
Figure S5. Physical properties, scaffold and chemotype analysis of screened fragment libraries.



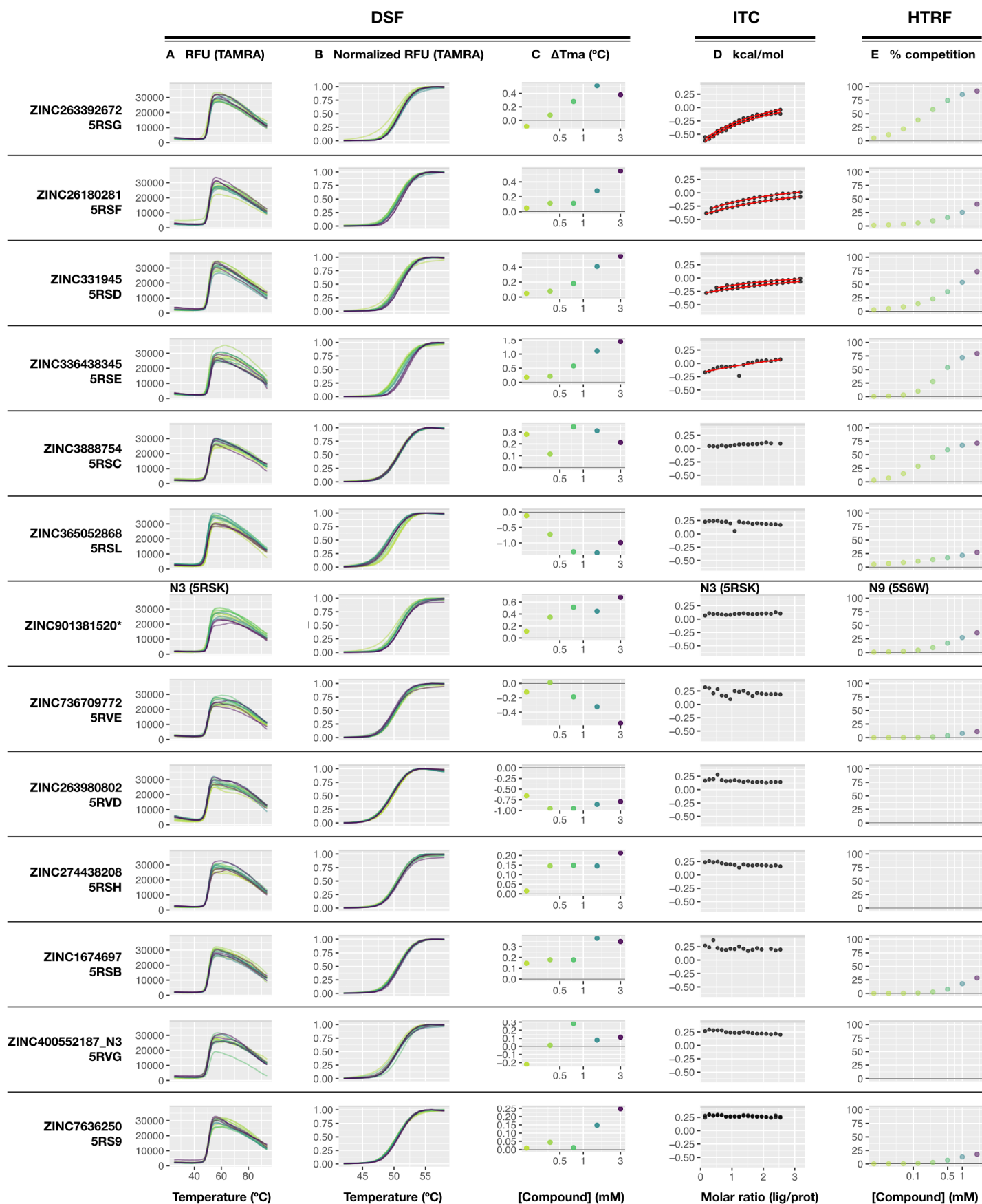
**Figure S6. Overview of fragment binding to protomer A (white surface) and protomer B (blue surface) of the P<sub>4</sub><sub>3</sub> crystals.**



**Figure S7. Additional soaking hits from docking and adenine-N3 vs -N9-alkylated isomers.**



**Figure S8. Mac1 subsites compared to the adenine binding site in kinases and the oxyanion binding site in carboxylesterases.** **A)** Key features of the Mac1 adenine subsite are illustrated by the structure of ZINC26180281 (PDB 5RSF). Hydrogen bonds are formed between the C6-amine of the adenine scaffold and the backbone nitrogen of Ile23, and between N1 of the adenine scaffold and the side-chain carboxylate of Asp22. The C2 amine of ZINC26180281 forms a non-canonical H-bond to the backbone carbonyl oxygen of Ala154. **B)** Adenine recognition is similar in the pseudokinase domain of JAK2, however, the C6-amine forms an H-bond to a backbone carbonyl oxygen rather than a side-chain carboxylate. Adenine binding occurs at the hinge residues that connect the N- and C-terminal lobes of the catalytic domain. Interactions that mimic adenine binding to the hinge residues are conserved in the majority of kinase inhibitors (57). Like ZINC2618028, kinase inhibitors exploit non-canonical H-bonds. The 1,2,4-triazole derived inhibitor shown in **(C)** forms an H-bond to the backbone carbonyl oxygen of Lys630. **D)** The fragment screens against Mac1 identified 47 oxyanions binding to the backbone nitrogens of Phe156 and Asp157. A comparable oxyanion recognition motif is present in acetylcholinesterase (AChE) **(E)**. In AChE, this motif stabilizes negative charge on the oxyanion transition state. **F)** Trifluoroacetic acid (TFA), present as a counter ion for ZINC3860798, was clearly defined in PanDDA event maps binding to the oxyanion subsite. TFA was also observed binding to the oxyanion subsite for fragments ZINC35185198 and ZINC51658946. The docking fragment ZINC263392672 also contained TFA, but no TFA was observed in the oxyanion subsite (PDB 5RSG).



**Figure S9. Comparison of DSF, HTRF, and ITC results for compounds tested in all assays.** **A)** Raw, un-normalized DSF data for the full measured temperature range (25 - 94°C) demonstrates the absence of confounding changes in curve shape for all compounds. **B)** Normalized raw DSF data, enlarged to visualize compound-induced thermal shifts. **C)** Changes in  $T_{m_a}$  observed in the presence of fragments (0-3 mM fragment). **D)** Integrated heat peaks as a function of binding site saturation shown as black dots. The red line represents a non-linear least squares (NLLS) fit using a single-site binding model. **E)** Dose-response curves showing competition of the fragments with an ADPr-conjugated peptide for Mac1 binding. (\*) ZINC901381520\_N3 was tested in DSF and ITC, ZINC901381520\_N9 was tested in HTRF.