Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease

Alice Douangamath[1,2], Daren Fearon[1,2], Paul Gehrtz [1,5] Tobias Krojer[1,4] , Petra Lukacik[1,2,3], C. David Owen[1,2,3], Efrat Resnick[1,5], Claire Strain-Damerell[1,2,3], Anthony Aimon[2,3], Péter Ábrányi-Balogh[6], José Brandaõ-Neto[2], Anna Carbery[2], Gemma Davison[9], Alexandre Dias[2], Thomas D Downes[8], Louise Dunnett[2], Michael Fairhead[4], James D. Firth[8], S. Paul Jones[8], Aaron Keely[6], György M. Keserü[6], Hanna F Klein[8], Mathew P. Martin[7], Martin E. M. Noble[7], Peter O'Brien[8], Ailsa Powell[2], Rambabu Reddi[5], Rachael Skyner[2], Matthew Snee[2], Michael J. Waring[9], Conor Wild[2], Nir London[5,*], Frank von Delft[2,3,4,10,*] & Martin A. Walsh[2,3*]

[1] These authors contributed equally

[2]Diamond Light Source Ltd., Harwell Science and Innovation Campus, Didcot OX11 0QX, U.K.

[3] Research Complex at Harwell, Harwell Science and Innovation Campus, Didcot OX11 0FA, U.K.

[4] Structural Genomics Consortium, University of Oxford, Old Road Campus, Roosevelt Drive, Headington, OX3 7DQ, UK.

[5]Department of Organic Chemistry, Weizmann Institute of Science, Rehovot 7610001, Israel

[6]Medicinal Chemistry Research Group, Research Centre for Natural Sciences, Magyar tudósok krt. 2, H-1117 Hungary

[7] Cancer Research UK Drug Discovery Unit, Newcastle University Centre for Cancer, Paul O'Gorman Building, Medical School, Framlington Place, Newcastle University, Newcastle upon Tyne, NE2 4AD, U.K.

[8]Department of Chemistry, University of York, Heslington, York, YO10 5DD, UK

[9] Cancer Research UK Drug Discovery Unit, Newcastle University Centre for Cancer, Chemistry, School of Natural and Environmental Sciences, Bedson Building, Newcastle University, Newcastle upon Tyne, NE1 7RU, U.K.

[10] Department of Biochemistry, University of Johannesburg, Auckland Park 2006, South Africa.

[*]Joint Corresponding authors

Nir London – nir.london@weizmann.ac.il

Frank von Delft - frank.von-delft@diamond.ac.uk

Martin A. Walsh - martin.walsh@diamond.ac.uk

Summary

COVID-19, caused by SARS-CoV-2, lacks effective therapeutics. Additionally, no antiviral drugs or vaccines were developed against the closely related coronavirus, SARS-CoV-1 or MERS-CoV, despite previous zoonotic outbreaks. To identify starting points for such therapeutics, we performed a large-scale screen of electrophile and non-covalent fragments through a combined mass spectrometry and X-ray approach against the SARS-CoV-2 main protease, one of two cysteine viral proteases essential for viral replication. Our crystallographic screen identified 71 hits that span the entire active site, as well as 3 hits at the dimer interface. These structures reveal routes to rapidly develop more potent inhibitors through merging of covalent and non-covalent fragment hits; one series of low-reactivity, tractable covalent fragments was progressed to discover improved binders. These combined hits offer unprecedented structural and reactivity information for on-going structure-based drug design against SARS-CoV-2 main protease.

Introduction

A novel coronavirus, SARS-CoV-2, the causative agent of COVID-19 (Wu et al., 2020, Kucharski et al., 2020, Zhu et al., 2020), has resulted in over one million confirmed cases and in excess of 300,000 deaths across 188 countries as of mid-May 2020 (Dong et al., 2020). SARS-CoV-2 is the third zoonotic coronavirus outbreak after the emergence of SARS-CoV-1 in 2002 and the Middle East Respiratory Syndrome (MERS-CoV) in 2012 (Bermingham et al., 2012, Kuiken et al., 2003, Zaki et al., 2012). SARS-CoV-2 is a large enveloped, positive sense, single stranded RNA Betacoronavirus. The viral RNA encodes two open reading frames that, through ribosome frame-shifting, generates two polyproteins pp1a and pp1ab (Bredenbeek et al., 1990). These polyproteins produce most of the proteins of the replicase-transcriptase complex (Thiel et al., 2003). The polyproteins are processed by two viral cysteine proteases: a Papain-like protease (PL^{pro}) which cleaves three sites, releasing non-structural proteins nsp1-3 and a 3C-like protease, also referred to as the main protease (M^{pro}), that cleaves at 11 sites to release non-structural proteins (nsp4-16). These non-structural proteins form the replicase complex responsible for replication and transcription of the viral genome and have led to M^{pro} and PL^{Pro} being the primary targets for antiviral drug development (Hilgenfeld, 2014).

Structural studies have played a key role in drug development and were quickly applied during the first coronavirus outbreak. Early work by the Hilgenfeld group facilitated targeting the M^{pro} of coronarviruses (Hilgenfeld, 2014), and this was accelerated during the 2002 SARS-CoV-1 outbreak, leading to crystal structures of SARS-CoV-1 M^{pro} and inhibitor complexes (Ghosh et al., 2007, Verschueren et al., 2008, Yang et al., 2005, Yang et al., 2003). Coronavirus M^{pro} active sites are well conserved (Anand et al., 2003, Hegyi and Ziebuhr, 2002, Stadler et al., 2003, Xue et al., 2008, Yang et al., 2005, Zhang et al., 2020b) and those of

enteroviruses ($3C^{pro}$) are functionally similar, which has led to efforts to develop broad spectrum antivirals. The most successful have been peptidomimetic α-ketoamide inhibitors (Zhang et al., 2020a), which have been used to derive a potent α-ketoamide inhibitor that may lead to a successful antiviral (Zhang et al., 2020b).

To date, no drugs targeting SARS-CoV-2 have been verified by clinical trials and treatments are limited to those targeting disease symptoms. To contribute to future therapeutic possibilities, we approached the SARS-CoV-2 M^{pro} as a target for high throughput drug discovery using a fragment-based approach (Thomas et al., 2019). We screened against over 1250 unique fragments leading to the identification of 74 high value fragment hits, including 23 non-covalent and 48 covalent hits in the active site, and 3 hits at the vital dimerization interface. Here, these data are detailed along with potential ways forward for rapid follow-up design of improved, more potent, compounds.

Results

M^{pro} crystallizes in a ligand-free form that diffracts to near-atomic resolution

We report the apo structure of SARS-CoV-2 M^{pro} with data to 1.25 Å. The construct we crystallised has native residues at both N- and C--terminals, without cloning truncations or appendages which could otherwise interfere with fragment binding. Electron density is present for all residues, including 26 alternate conformations, many of which were absent in previous lower resolution crystal structures. The protein crystallised with a single protein polypeptide in the asymmetric unit, and the catalytic dimer provided by a symmetry-related molecule. The structure aligns closely with the M^{pro} structures from SARS-CoV-1 and MERS (rmsd of 0.52 Å and 0.97 Å respectively). The active site is sandwiched between two β-barrel domains, I (residue 10-99) and II (residue 100-182) (Figure 1A). Domain III (residue 198-306), forms a bundle of alpha helices and is proposed to regulate dimerization (Shi and Song, 2006). The C-terminal residues, Cys300-Gln306, wrap against Domain II. However, the C terminal displays a degree of flexibility and wraps around domain III in the N3 inhibitor complex (Shi and Song, 2006) (PDB ID 6LU7). His41 and Cys145 comprise the catalytic dyad and dimerisation completes the active site by bringing Ser1 of the second dimer protomer into proximity with Glu166 (Figure 1B). This aids formation of the substrate specificity pocket and the oxyanion hole (Hilgenfeld, 2014). Subsites have previously been identified in the active site based on interactions with peptide-based inhibitors and are shown in figure 1B (Jin et al., 2020, Zhang et al., 2020b). Comparisons with peptide-based inhibitor complexes (Jin et al., 2020, Zhang et al., 2020b) suggest a degree of active site plasticity. In particular, the C-alphas of Met49, Pro168, Gln189 respectively show movements of 2.8 Å, 1.4 Å, and 1.2 Å in comparison to the α-ketoamide inhibitor bound M^{pro} structure (Zhang et al., 2020b) (PDB ID 6y2f, Figure 1B).

The crystal form is well-suited for crystallographic fragment screening: although the percentage of solvent is very low for a protein crystal, approximately 20%, nevertheless clear channels are present that allow access to the active site through diffusion. Moreover, the

tight packing and strong innate diffraction mean crystals are resistant to lattice disruption and degradation of diffraction by DMSO solvent when adding solubilised fragments to the crystallization drop.

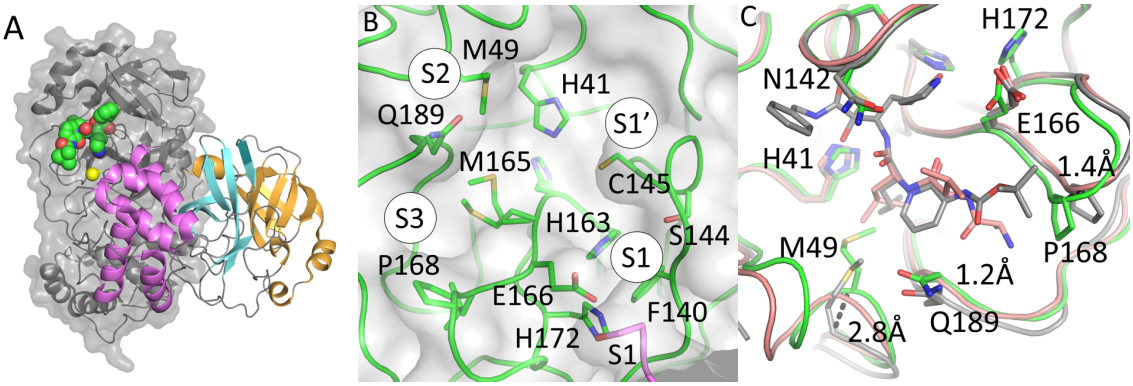


Figure 1. The crystal structure of ligand free M^{pro} is amenable to X-ray fragment screening. A. cartoon representation of the M^{pro} dimer. The nearmost monomer is shown with secondary structure features coloured to demarcate domains I, II, and III, in orange, cyan, and violet respectively. The active site of the rear monomer is indicated by the presence of a peptide-based inhibitor in green generated by aligning the ligand-free structure with pdb 6Y2F (Zhang et al., 2020b). A yellow sphere indicates Ser1 from the dimer partner that completes the active site. **B.** Residues of the active site are labelled, and subsites involved in ligand binding are shown with circles. **C.** Active site plasticity is observed when comparing apo structure to peptide inhibitor bound structures (green – Apo, grey – 6Y2F, pink 6LU7). Displacement distances associated with loop movements are indicated.

Combined MS and crystallographic fragment screens reveal new binders of M^{pro}

Cysteine proteases are attractive targets for covalent inhibitors. To identify covalent starting points, we screened our previously described library of ~1000 mild electrophilic fragments (Resnick et al., 2019) against M^{pro} using intact protein mass spectrometry. Standard conditions of 200 µM per electrophile for 24 hours at 4 °C did not allow discrimination between hits. Screening at more stringent conditions (5 µM per electrophile; 1.5 hours; 25°C) resulted in 8.5% of the library labelling above 30% of protein (Table S1a). These hits revealed common motifs, and we focused on compounds which offer promising starting points.

Compounds containing *N*-chloroacetyl-*N*'-sulfonamido-piperazine or *N*--chloroacetylaniline motifs were frequent hitters. Such compounds can be highly reactive. Therefore, we chose series members with relatively low reactivity for follow up crystallization attempts. For another series of hit compounds, containing a *N*-chloroacetyl piperidinyl-4-carboxamide motif (Table S2) which displays lower reactivity and were not frequent hitters in previous screens, we attempted crystallization despite their absence of labelling in the stringent conditions.

While mild electrophilic fragments are ideal for probing the binding properties around the active site cysteine, their small size prevents extensive exploration of the substrate binding pocket. We performed an additional crystallographic fragment screen to exhaustively probe the M^{pro} active site, and to find opportunities for fragment merging or growing. The 68 electrophile fragment hits were added to crystals along with a total of 1176 unique fragments

from 7 libraries (Table S3). Non-covalent fragments were soaked (Collins et al., 2017), whereas electrophile fragments were both soaked and co-crystallized as previously described (Resnick et al., 2019), to ensure that as many of the mass spectrometry hits as possible were structurally observed. A total of 1742 soaking and 1139 co-crystallization experiments resulted in 1877 mounted crystals. While some fragments either destroyed the crystals or their diffraction, 1638 datasets with a resolution better than 2.8 Å were collected. The best crystals diffracted to better than 1.4 Å, but diffraction to 1.8 Å was more typical, and no datasets worse than 2.8 Å were included in analysis (Figure S2). We identified 96 fragment hits using the PanDDA method (Pearce et al., 2017), all of which were deposited in the Protein Data Bank (Table S2), but also immediately released through the Diamond Light Source website (https://www.diamond.ac.uk/covid-19.html), along with all protocols and experimental details. A timeline of experiments is shown in figure 2.

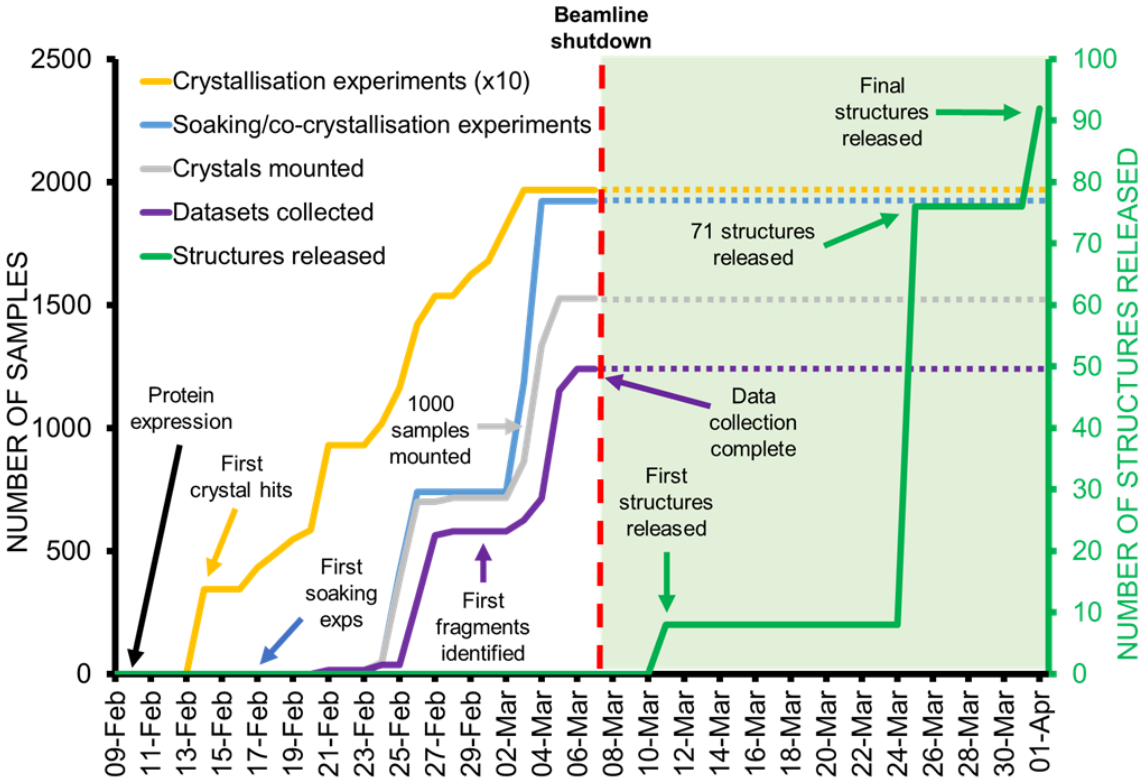


Figure 2. Timeline of crystallographic fragment screen.

Non-covalent fragment hits reveal multiple targetable sub-sites in the active site

This unusually large screen identified 23 structurally diverse fragments that bind non-covalently and extensively sample features of the M^{pro} active site and its specificity pockets/subsites (figure 1), along with 3 hits exploring the dimer interface.

Active-site fragments

Eight fragments were identified that bind in the S1 subsite and frequently form interactions with the side chains of the key residues His163, through a pyridine ring or similar nitrogen containing heterocycle, and Glu166 through a carbonyl group in an amide or urea moiety (Figure 3). Several also reach across into the S2 subsite.

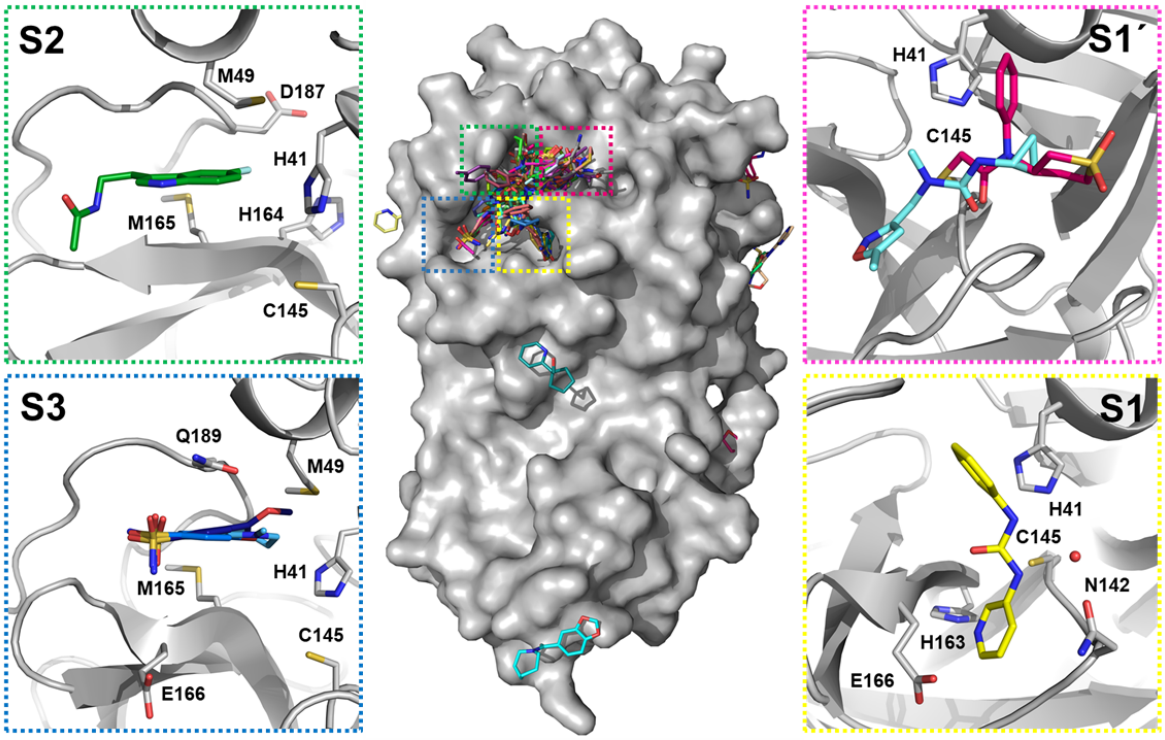


Figure 3. Bound fragments sample the active site comprehensively. The central surface representation is of the M^{pro} monomer with all fragment hits shown as sticks, and active site subsites highlighted by coloured boxes. Each subsite expanded along with a selection of hits to demonstrate common features and interactions. **S1**: Z44592329 (x0434); **S1´**: Z369936976 (x0397) in aquamarine and PCM-0102372 (x1311) in magenta bound to active site cysteine; **S2**: Z1220452176 (x0104); **S3**: Overlay of Z18197050 (x0161), Z1367324110 (x0195) and NCL-00023830 (x0946).

Subsite S2 has previously demonstrated greater flexibility in comparison to the other subsites, adapting to smaller substituents in peptide-based inhibitors but with a preference for leucine or other hydrophobic residues (Zhang et al., 2020b). Many fragments bound at this location, which we termed the “aromatic wheel” because of a consistent motif of an aromatic ring forming hydrophobic interactions with Met49 or π- π stacking with His41, with groups variously placed in 4 axial directions. Particularly notable is the vector into the small pocket between His164, Met165 and Asp187, exploited by three of the fragments (Z1220452176 (x0104), Z219104216 (x0305) and Z509756472 (x1249)) with fluoro and cyano substituents (Figure 3).

Of the four fragments exploring subsite S3, three contain an aromatic ring with a sulfonamide group forming hydrogen bonds with Gln189 and pointing out of the active site towards the solvent interface (Figure 3). These hits have expansion vectors suitable for exploiting the same His164/Met165/Asp187 pocket mentioned above.

The experiment revealed one notable conformational variation, which was exploited by one fragment only (Z369936976 (x0397); Figure 4): a change in the sidechains of the key catalytic residues His41, Cys145 alters the size and shape of subsite S1’ and thus the link to subsite S1. This allows the fragment to bind, uniquely, to both S1 and S1’. In S1, the isoxazole nitrogen hydrogen-bonds to His163, an interaction that features in several other hits; and in S1’, the

cyclopropyl group occupies the region sampled by the covalent fragments. Notably, the N-methyl group offers a vector to access the S2 and S3 subsites.

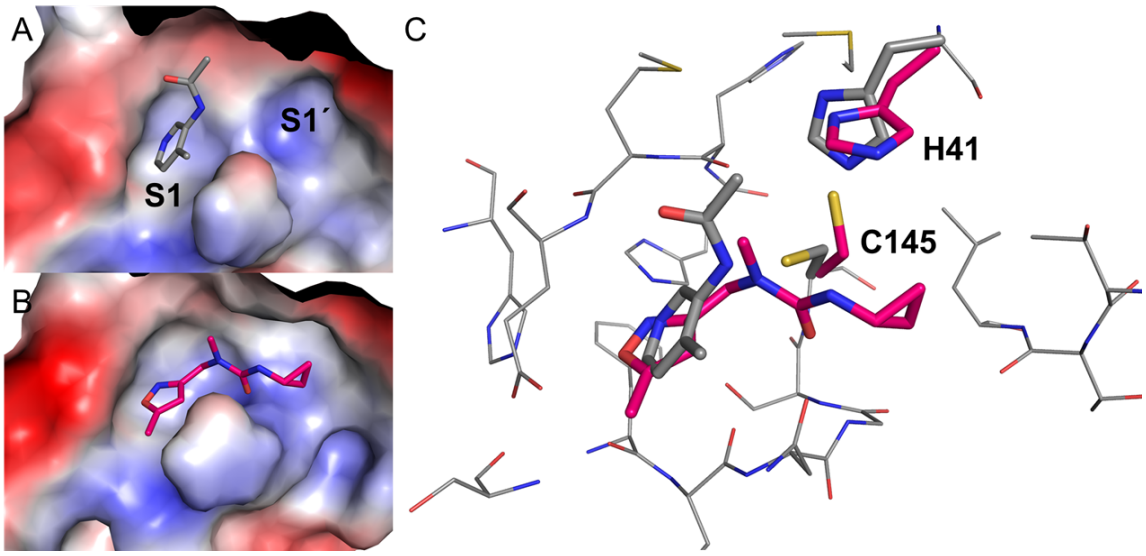


Figure 4. Plasticity of S1' is revealed by fragment Z369936976 (x0397). Comparing the electrostatic surfaces of Z1129283193 (x0107) (A), the most commonly observed conformation, with that of Z369936976 (x0397) (B) shows how the shape of S1 and S1' can change. C: Sidechain movement of catalytic residues Cys145 and His41 upon binding of Z369936976 (x0397, magenta) compared to Z1129283193 (x0197, grey).

Dimer interface fragments

It is established that the biological unit for similar viral proteases, such as the SARS-CoV-1 protease is a dimer (Chou et al., 2004), and that mutations at the dimer interface can disrupt proteases activity (Chen et al., 2008, Hsu et al., 2005) even at long range (Barrila et al., 2006). Thus, compounds that interfere with dimerization might serve as quasi-allosteric inhibitors of protease activity. In this study three compounds bound at the dimer interface.

Fragment Z1849009686 (x1086; Figure 5A) binds in a hydrophobic pocket formed by the sidechains of Met6, Phe8, Arg298 and Val303. It also mediates two hydrogen bonds to the sidechain of Gln127 and the backbone of Met6. Its binding site is less than 7 Å away from Ser139, whose mutation to alanine in SARS-CoV-1 protease reduced both dimerization and protease activity by about 50% (Chen et al., 2008, Hu et al., 2009). Z264347221 (x1187, Figure 5B) binds similarly in a hydrophobic pocket made by Met6, Phe8 and Arg298 in one of the protomers, extending across the dimer interface to interact with Ser123, Tyr118 and Leu141 of the second protomer, including hydrogen bonds with the sidechain and backbone of Ser123. Finally, POB0073 (x0887; York 3D library; Figure 5C), binds only 4 Å from Gly2 at the dimer interface and is encased between Lys137 and Val171 of one protomer and Gly2, Arg4, Phe3, Lys5 and Leu282 of the second, including two hydrogen bonds with the backbone of Phe3.

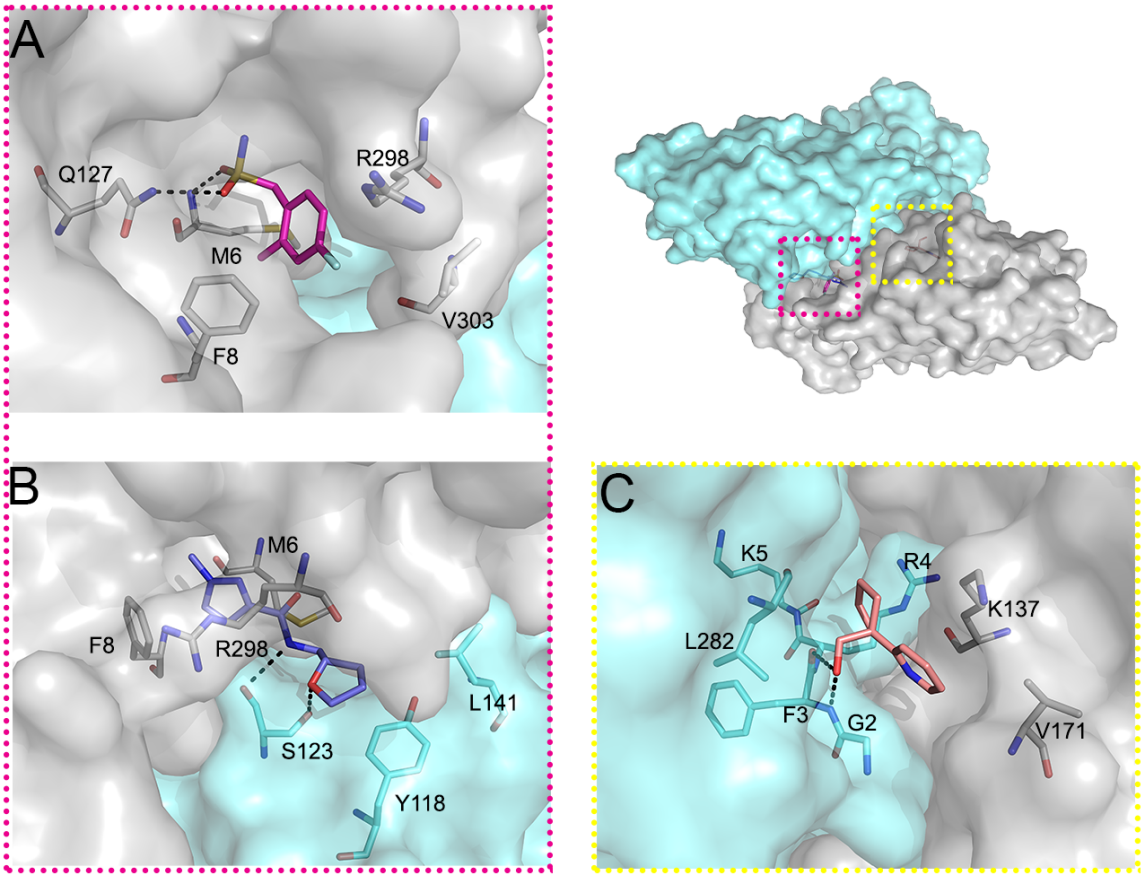


Figure 5. Fragments at dimer interface indicate opportunities for allosteric modulation. The overview shows the surface of the M^{pro} dimer, the protomers in grey and cyan. Fragments and surrounding residues are shown as sticks and hydrogen bonds in dashed black lines. A. Z1849009686 (x1086). B. Z264347221 (x1187). C. POB0073 (x0887).

Covalent fragment hits reveal several tractable series

The screen further yielded 48 structures of fragments covalently bound to the nucleophilic active site Cys145, and sample subsite S1'. The majority (44) fall into series explored in the mass spectrometry experiment and the remainder came from other libraries.

Electrophile fragments

In all structures with bound electrophiles, the *N*-chloroacetyl carbonyl oxygen atom forms either two or three hydrogen bonds with the backbone amide hydrogens of Gly143, Ser144 or Cys145 (Figure 6 A-C). All three compounds containing the *N*-chloroacetyl piperidinyl-4-carboxamide motif (Figure 6A) adopt a similar binding mode pointing towards the S2 pocket, and one (PCM-0102389, x1358) is able to form an additional hydrogen bond with the side chain of Asn142.

Compounds with the *N*-chloroacetyl-*N*'-sulfonamido-piperazine motif (Figure 6B) adopt a bent shape, pointing towards the S2 pocket where appropriate space-filling substituents are attached to the phenyl moiety (PCM-0102353 (x1336) and PCM-0102395 (x0774)); otherwise, they point towards the solvent. Most of the latter 8 structures feature a halophenyl moiety

which resides closely to Asn142, hinting at weak halogen-mediated interactions (Kuhn et al., 2019).

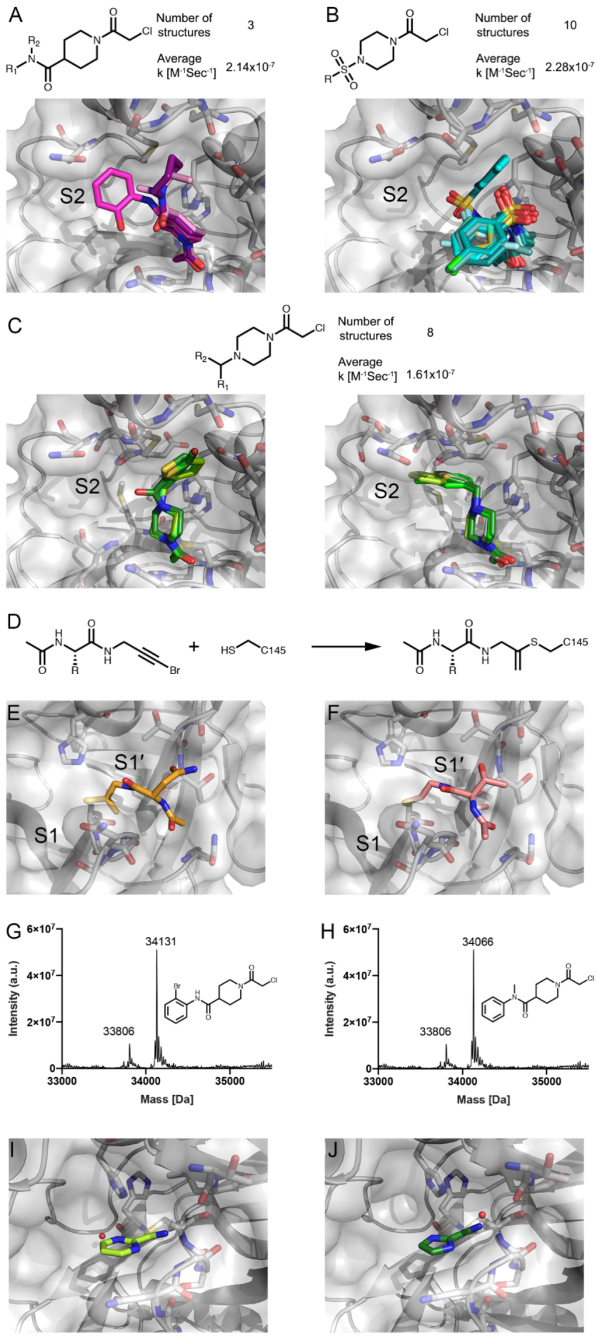


Figure 6. Covalent fragments are anchored at Cys145 and sample different regions of the orthosteric M^{pro} binding pocket. A. Fragments containing *N*-chloroacetyl piperidinyl-4-carboxamide motif. B. Fragments containing *N*-chloroacetyl-*N*'-sulfonamido-piperazine motif. C. Fragments containing *N*-chloroacetyl-*N*'-carboxamido- and *N*-chloroacetyl-*N*'-heterobenzyl-piperazine in two binding modes. D: Reaction schema of the unexpected covalent modification to Cys145 by PepLites hits. E. Threonine PepLite (NCL-00025058 (x0978)) bound covalently to active site cysteine. F: Asparagine PepLite (NCL-00025412 (x0981)) bound to active site cysteine. Labelling of M^{pro} by 2nd generation compounds proven by intact protein LC-MS: G. labelling by PG-COV-35; H. labelling by PG-COV-34. Covalently bound cyclic electrophiles: I. Cov_HetLib 030 (x2097) and J. Cov_HetLib 053 (x2119).

Eight compounds with a *N*-chloroacetyl-*N*'-carboxamido- and *N*-chloroacetyl-*N*'-heterobenzyl-piperazine motif crystallized in one binding mode with respect to the piperazinyl moiety (Figure 6C) (with one exception, PCM-0102287 (x0830)). Two structures (PCM-0102277 (x1334), PCM-0102169 (x1385)) with a 5-halothiophen-2-ylmethylene moiety exploit lipophilic parts of S2, which is also recapitulated by the thiophenyl moiety in an analogous carboxamide (PCM-0102306 (x1412)). The other five structures point mainly to S2, offering an accessible growth vector towards the nearby S3 pocket.

A series of compounds containing a N-chloroacetyl piperidinyl-4-carboxamide motif showed promising binding modes. To follow up on these compounds we performed a rapid second-generation compound synthesis. Derivatives of this chemotype were accessible in mg-scale by reaction of N-chloroacetyl piperidine-4-carbonyl chloride with various in-house amines, preferably carrying a chromophore to ease purification. These new compounds were tested by intact protein mass-spectrometry to assess protein labelling (5 uM compound; 1.5h incubation, RT; Table S1b). Amides derived from non-polar amines mostly outcompeted their polar counterparts, hinting at a targetable lipophilic sub-region in this direction. The two amides with the highest labelling PG-COV-35 and PG-COV-34 (figure 6G,H) highlight the potential for further synthetic derivatization by amide N-alkylation or cross-coupling, respectively.

PepLites

The screen revealed unexpected covalent warheads from the series of 3-bromoprop-2-yn-1-yl amides of N-acylamino acids. Colloquially termed PepLites (Noble and Waring), this library was developed to map non-covalent interactions of amino acid sidechains in protein-protein interaction hotspots, with the acetylene bromine intended, as for FragLites (Bauman et al., 2016, Wood et al., 2019), as detection tag by anomalous dispersion in X-ray crystallography. However, bromoalkynes can also act as covalent traps for activated cysteine thiols (Mons et al., 2019) (figure 6D).

Two PepLites, containing threonine (NCL-00025058 (x0978)) and asparagine (NCL00025412 (x0981)) bound covalently to the active site cysteine (Cys145), forming a thioenolether via C-2 addition with loss of bromine (Figure 6E,F). The covalent linkage was unexpected and evidently the result of significant non-covalent interactions, specific to these two PepLites, that position the electrophile group for nucleophilic attack. We note the side-chains make hydrogen-bonding interactions with various backbone NH and O atoms of Thr26 and Thr24; in the case of threonine, it was the minor 2R,3R diastereomer (corresponding to D-allothreonine) that bound. The only other PepLite observed (tyrosine, NCL-00024905 (x0967)) bound non-covalently to a different subsite.

The highlighted structure activity relationships is important for further optimisation. Bromoalkynes have intrinsic thiol reactivity that is lower than that of established acrylamide-based covalent inhibitors (Mons et al., 2019), which is in general desirable. The geometry of the alkyne and its binding mode also suggest that it could be replaced by reversible covalent groups such as nitriles, which would be guided by the same non-covalent interactions but are better established as cysteine protease inhibitors.

Heterocyclic electrophiles

Two covalent hits (2-cyano pyrimidine (Cov_HetLib 030 (x2097)) and 2cyano-imidazole (Cov_HetLib 053 (x2119) came from a library of small heterocyclic electrophiles (Keeley et al., 2019). These are essentially covalent MiniFrags (O'Reilly et al., 2019), comprising five and six-membered nitrogen containing heterocycles with electron-withdrawing character that activates small electrophilic substituents (halogenes, acetyl, vinyl and nitrile groups).

Both hits bound to Cys145 through an imine (Figure 6I,J), positioned by a local hydrogen bond network involving imine and heterocyclic N atoms. One of these free amines provides an immediate growth vector towards to catalytic pocket. The compounds have reasonable stability in water (Keeley et al., 2018) and limited reactivity against GSH (t1/2= 2.2 and 52.3 h, respectively), well above suggested reactivity limits (Fuller et al., 2016). They are also inactive against various covalent targets (HDAC8, MAO-A, MAO-B, MurA) and benchmark proteins.

Discussion

The data presented herein provides many clear routes to developing potent inhibitors against SARS-CoV-2. The bound fragments comprehensively sample all subsites of the active site revealing diverse expansion vectors, and the electrophiles provide extensive, systematic as well as serendipitous, data for designing covalent compounds.

It is widely accepted that new small molecule drugs cannot be developed fast enough to help against COVID-19. Nevertheless, as the pandemic threatens to remain a long-term problem and vaccine candidates do not promise complete and lasting protection, antiviral molecules will remain an important line of defence. Such compounds will also be needed to fight future pandemics (Hilgenfeld, 2014). Our data will accelerate such efforts: therapeutically, through design of new molecules and to inform ongoing efforts at repurposing existing drugs; and for research, through development of probe molecules (Arrowsmith et al., 2015) to understand viral biology. One example is the observation that fragment Z1220452176 (x0104) is a close analogue of melatonin, although in this case, it is unlikely that melatonin mediates direct antiviral activity through inhibition of M^{pro}, given its low molecular weight; nevertheless, melatonin is currently in clinical trials to assess its immune-regulatory effects on COVID19 (Clinicaltrials.gov identifier NCT04353128).

In line with the urgency, results were made available online immediately for download. Additionally, since exploring 3D data requires specialised tools (Ferla et al., 2020, Lee et al., 2011), hits were made accessible on the Fragalysis webtool (https://fragalysis.diamond.ac.uk) that allows easy exploration of the hits in interactive 3D.

We have previously demonstrated the benefits of merging covalent and non-covalent fragments to make dramatic improvements in potency (Resnick et al., 2019). Our dataset offers numerous opportunities and some conservative examples are shown in figure 7. These can be expected to result in potent M^{pro} binders and compound synthesis is ongoing.

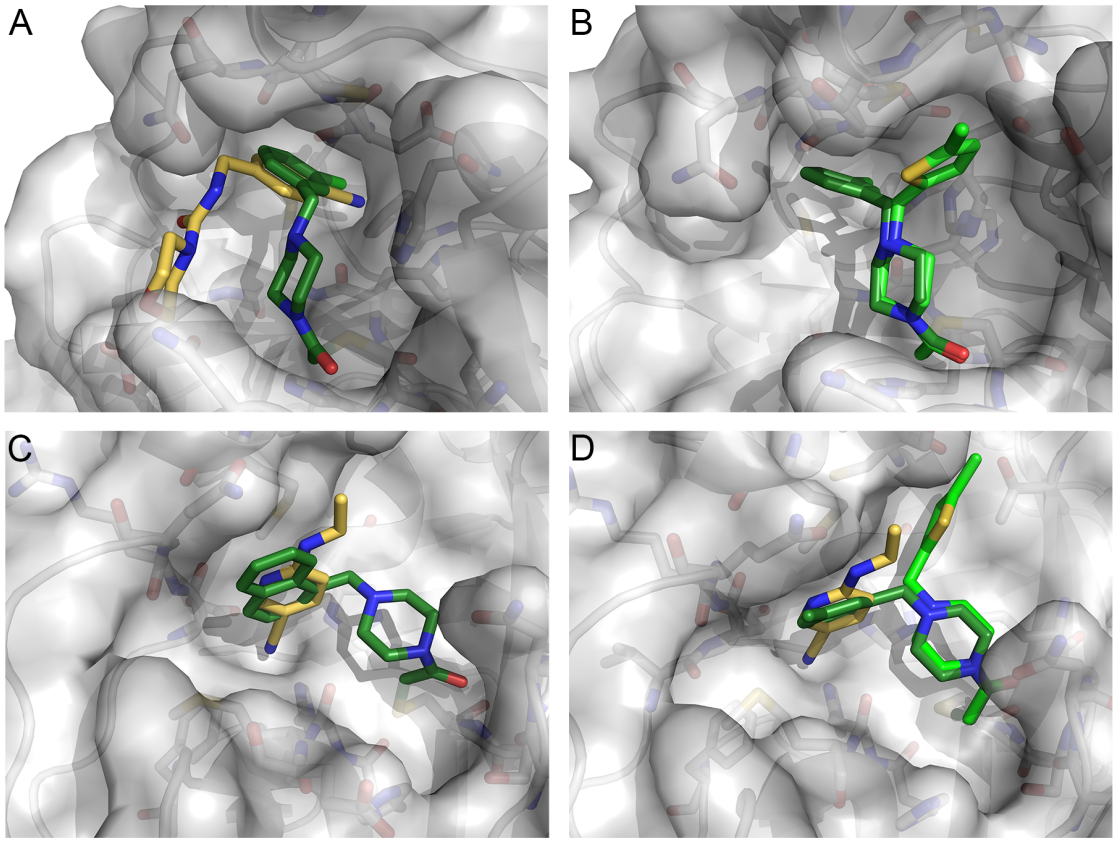


Figure 7. Fragment merging opportunities can be directly inferred from many hits. Covalently-bound fragments are in green shades, and non-covalent fragments in yellow A. Overlay of Z509756472/x1249 and PCM-0102269/x0770. B. Overlay of PCM-0102277/x1334 and PCM-0102269/x0770. C. Overlay of PCM-0102287/x0830 and Z219104216/x0305. D. Overlay of PCM-0102340/x0692, PCM-0102277/x1334 and Z219104216/x0305.

Collectively, the covalent hits provide rational routes to inhibitors of low reactivity and high selectivity. Rationally designed covalent drugs are gaining traction, with many recent FDA approvals (Singh et al., 2011, Bauer, 2015). Their design is based on very potent reversible binding, that allows precise orientation of a low reactivity electrophile, so that formation of the covalent bond is reliant on binding site specificity, with minimal off-targets. (De Cesco et al., 2017, Zhang et al., 2019, Baillie, 2016). For this reason, covalent inhibitors are expunged from high-throughput screening libraries and are typically considered as PAINS compounds (Sirois et al., 2005, Baell and Nissink, 2018, Baell and Holloway, 2010). The challenge of tuning reactivity, and the danger of reactivity-based artefacts, are likely to be particularly marked for the highly reactive nucleophiles such as the catalytic cysteine of many proteases. This is evidenced by the very high hit-rate we saw in our preliminary electrophiles screen in which more than 150 fragments labelled M^{pro} by >50%. Robust characterization of the fragments' reactivity (Resnick et al., 2019), and continuous evaluation of general thiol reactivity in the selection of lead series and during hit-to-lead optimization can address this challenge.

The scale of this experiment, particularly the diversity of libraries and density of results, likely sets a new benchmark for ensuring a crystal-based fragment screen accelerates progression of hits. Even cursory inspection of the fragment structures indicates a very large "merge space", i.e. the collection of compounds that can be designed directly from spatial juxtapositions of fragments. Such merges, which can be made to populate all four subsites, might achieve potency synergistically, because the observed interactions can be assumed to be in near-optimal configurations, given how few there are per fragment. A thorough exploration of merge space might be best achieved formulaically, using computational workflows that additionally filter undesirable molecular properties, assess synthetic tractability and predict binding affinity. However, such integrated approaches are not currently available in the public domain. We hope this dataset will help spur their development and testing.

Another promising effort to explore the potential of this premise is the COVID Moonshot project (https://covid.postera.ai/covid), where a selection of merges will be experimentally tested, with data promptly made public. We trust that this resource will enable the development of many new tools, approaches and ultimately viable treatment candidates for COVID19.

Materials and Methods

Protein Expression: Multiple transformant colonies were used to inoculate a starter culture supplemented with 100 µg/ml Carbenicillin. The culture was then grown to log phase for approximately 8 h. 10 ml of the starter culture was used to inoculate 1 litre of Auto Induction Medium supplemented with 10 ml of glycerol and 100 µg/ml Carbenicillin. The cultures were grown at 37 °C, 200 rpm for 5 h then switched to 18 °C, 200 rpm for 10 h. The cells were harvested by centrifugation and stored at -80 °C

Protein purification: Cells were resuspended in 50 mM Tris pH 8, 300 mM NaCl, 10 mM Imidazole, 0.03 µg/ml Benzonase. The cells were disrupted on a high-pressure homogeniser (3 passes, 30 kpsi, 4 °C). The lysate was clarified by centrifugation at 50,000 g. The supernatant was then applied to a Nickel-NTA gravity column and washed and eluted with 50 mM Tris pH 8, 300 mM NaCl, and 25-500 mM imidazole pH 8. N-terminal His tagged HRV 3C Protease was then added to the eluted protein at 1:10 w/w ratio. The mixture was then dialysed overnight at 4 °C against 50 mM Tris pH 8, 300 mM NaCl, 1 mM TCEP. The following day, the HRV 3C protease and other impurities were removed from the cleaved target protein by reverse Nickel-NTA. The relevant fractions were concentrated and applied to an S200 16/60 gel filtration column equilibrated in 20 mM Hepes pH 7.5, 50 mM NaCl buffer. The protein was concentrated to 30 mg/ml using a 10 kDa MWCO centrifugal filter device.

Crystallisation and structure determination: Protein was thawed and diluted to 5 mg/ml using 20 mM Hepes pH 7.5, 50 mM NaCl. The sample was centrifuged at 100 000 g for 15 minutes. Initial hits were found in well F2 of the Proplex crystallisation screen, 0.2 M LiCl, 0.1M Tris pH 8, 20% PEG 8K. These crystals were used to prepare a seed stock by crushing the proteins with a pipette tip, suspending in reservoir solution and vortexing for 60 s in the reservoir solution with approximately 10 glass beads (1.0mm diameter, BioSpec products).

Adding DMSO to the protein solution to a concentration of 5% and performing microseed matrix screening, many new crystallisation hits were discovered in commercial crystallisation screens. Following optimisation, the final crystallisation condition was 11% PEG 4K, 6% DMSO, 0.1M MES pH 6.7 with a seed stock dilution of 1/640. The seeds were prepared from crystals grown in the final crystallisation condition. The drop ratios were 0.15 µl protein, 0.3 µl reservoir solution, 0.05 µl seed stock. Crystals were grown using the sitting drop vapor diffusion method at 20 °C and appeared within 24 hours.

Initial diffraction data was collected on beamline I04 at Diamond Light Source on a crystal grown in 0.1 M MES pH 6.5, 5% PEG6K, cryoprotected using 30% PEG400. Data were processed using Dials (Winter et al., 2018) via Xia2 (Winter et al., 2013). The dataset was phased with the SARS-CoV-2 M^{pro} in complex with the N9 inhibitor crystal structure (PDB:6LU7) using Molrep in CCP4i2. Further datasets were collected on I04-1 at Diamond Light Source on crystals grown using the 0.1 M MES pH 6.5, 15% PEG4K, 5% DMSO condition. To create a high-resolution dataset, datasets from 7 crystals were scaled and merged using Aimless (Evans and Murshudov, 2013). Crystal structures were manually rebuilt in Coot (Emsley et al., 2010) and refined using Refmac (Murshudov et al., 2011) and Buster (Bricogne et al., 2017). This structure is deposited in the PDB under ID 6YB7.

Electrophile fragment LC/MS screen: 2 µM M^{pro} was incubated in 50 mM Tris pH 8 300 mM NaCl for 1.5 hours at 25 °C. For initial electrophile fragment library screen, 30 µl protein with pools of 4-5 electrophile fragments, 7.5 nL each from 20 mM DMSO stocks and for other runs 50 µl protein with 0.5 µl compounds from 0.5 mM DMSO stocks. The reaction was quenched by adding formic acid to 0.4% final concentration. The LC/MS runs were performed on a Waters ACUITY UPLC class H instrument, in positive ion mode using electrospray ionization. UPLC separation used a C4 column (300 Å, 1.7 µm, 21 mm × 100 mm). The column was held at 40 °C and the autosampler at 10 °C. Mobile solution A was 0.1% formic acid in water, and mobile phase B was 0.1% formic acid in acetonitrile. The run flow was 0.4 mL/min with gradient 20% B for 4 min, increasing linearly to 60% B for 2 min, holding at 60% B for 0.5 min, changing to 0% B in 0.5 min, and holding at 0% for 1 min. The mass data were collected on a Waters SQD2 detector with an m/z range of 2−3071.98 at a range of 1000−2000 m/z. Raw data were processed using openLYNX and deconvoluted using MaxEnt. Labelling assignment was performed as previously described (Resnick et al., 2019).

Fragment Screening: Fragments were soaked into crystals as previously described (Collins et al., 2017), by adding dissolved compound directly to the crystallisation drops using an ECHO liquid handler (final concentration 10% DMSO); drops were incubated for approximately 1 hour prior to mounting and flash freezing in liquid nitrogen. The following libraries were screened: the DSi-poised library (Enamine), a version of the poised library (Cox et al., 2016); a version of the MiniFrags library (O'Reilly et al., 2019) assembled in-house; the FragLites library (Wood et al., 2019); a library of shape-diverse 3D fragments (“York3D”) (Downes et al., 2020); heterocyclic electrophiles (Keeley et al., 2019); and the SpotFinder library (Bajusz and Keserü). All fragments were in 100% DMSO at varying stock concentrations, detailed at https://www.diamond.ac.uk/Instruments/Mx/Fragment-Screening/Fragment-Libraries.html).

Electrophile fragments identified by mass spectrometry were soaked by the same procedure as the other libraries, but in addition, they were also co-crystallised in the same crystallisation condition as for the apo structure. The protein was incubated with 10 to 20-fold excess compound (molar ratio) for approximately 1h prior to the addition of the seeds and reservoir solution (following Resnick *et al* (Resnick et al., 2019)).

Data were collected at the beamline I04-1 at 100K and processed to a resolution of approximately 1.8 Å using XDS (Kabsch, 2010) and either xia2 (Winter et al., 2013), autoPROC (Vonrhein et al., 2011) or DIALS (Winter et al., 2018). Further analysis was performed with XChemExplorer (Krojer et al., 2017): electron density maps were generated with Dimple (Keegan et al., 2015); ligand-binding events were identified using PanDDA (Pearce et al., 2017) (both the released version 0.2 and the pre-release development version (https://github.com/ConorFWild/pandda)); ligands were modelled into PanDDA-calculated event maps using Coot (Emsley et al., 2010); restraints were calculated with ACEDRG or GRADE (Long et al., 2017, Smart et al., 2010); and structures were refined with Refmac (Murshudov et al., 2011) and Buster (Bricogne et al., 2017). A more thorough description of the PanDDA analysis is provided in the supplementary information.

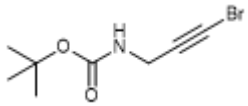
Coordinates, structure factors and PanDDA event maps for all data sets are deposited in the Protein Data Bank under group deposition ID G_1002135, G_1002151, G_1002152, G_1002153, G_1002156 and G_1002157. Data collection and refinement statistics are summarised in supplementary table 4. The ground-state structure and all corresponding datasets are deposited under PDB ID 5R8T.

Synthesis of *N*-chloroacetyl-piperidine-4-carboxamides *N*-chloroacetyl piperidine-4-carbonyl chloride was prepared as a stock solution in dry DCM under an atmosphere of N_2. Briefly, deprotecting *N*-Boc isonepecotic acid in 50% TFA in DCM (v/v) at RT for 2 h yielded the corresponding TFA salt after evaporation of all volatiles. The crude TFA salt was then re-dissolved in DCM, treated with Et_3N (2 equiv.), followed by the addition of chloroacetic anhydride (1 equiv.). The reaction mixture was stirred overnight at RT, washed with water, the organic phase dried over $MgSO_4$, filtered, and all volatiles removed by rotary evaporation. The crude N-chloroacetyl piperidine-4-carboxylic acid was refluxed in excess neat $SOCl_2$ (gas evolution and a colour change to red occurs) for 1 h, followed by removal of excess $SOCl_2$ in vacuum into a liquid nitrogen-cooled trap. The remaining residue was dried by rotary evaporation, placed under an atmosphere of nitrogen and dissolved in dry DCM to give a stock solution of approx. 0.489 M (based on theoretical yield over three steps), which was immediately used.

The corresponding amides were prepared by addition of the acid chloride (1 equiv.) as a DCM solution to the pertinent amines (1 equiv.) in presence of pyridine (1 equiv.) in DCM. Heterogeneous reaction mixtures were treated with a minimal amount of dry DMF to achieve full solubility. After stirring the reaction mixtures overnight, the solvents were removed in by rotary evaporation, re-dissolved in 50% aq. MeCN (and a minimal amount of DMSO to achieve higher solubility), followed by purification by (semi-)preparative RP-HPLC in mass-directed automatic mode or manually.

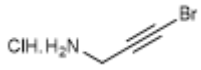
Synthesis of PepLites HATU (1.5 eq.), DIPEA (3.0 eq.) and the acid starting material (1.5 eq.) were dissolved in DMF (3 - 6 mL) and stirred together at room temperature for 10 min. 3-Bromoprop-2-yn-1-amine hydrochloride was added and the reaction mixture was stirred at 40 °C overnight. The reaction mixture was allowed to cool to room temperature, diluted with EtOAc or DCM and washed with saturated aqueous sodium bicarbonate solution, brine and water. The organic layer was dried over $MgSO_4$, filtered and evaporated to afford crude product. The crude product was then purified by either normal or reverse phase chromatography.

***tert*-Butyl (3-bromoprop-2-yn-1-yl)carbamate**



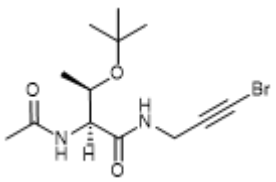
A solution of KOH (2.7 g, 48 mmol) in water (15 mL) was added dropwise to a solution of *N*-bocpropargylamine (3.0 g, 19 mmol) in MeOH (45 mL) at 0 °C under nitrogen. The resulting solution was stirred at 0 °C for 10 min then bromine (1.1 mL, 21 mmol) was added dropwise. The reaction mixture was allowed to warm to room temperature and was stirred at room temperature for 24 h. The reaction mixture was diluted with water and extracted with diethyl ether. The organic extracts were combined, dried over $MgSO_4$ and evaporated to afford crude product. The crude product was purified by flash silica chromatography, elution gradient 0 – 10% EtOAc in petroleum ether. Pure fractions were evaporated to dryness to afford *tert*-Butyl (3-bromoprop-2-yn-1-yl)carbamate (3.5 g, 79%) as a white solid. R_f = 0.34 (10% EtOAC in petroleum ether); m.p. 108 - 110 °C; IR ν_{max} (cm^{-1}) 3345, 2982, 2219, 2121, 2082; ^{1}H NMR (500 MHz, DMSO-*d*6) δ 1.39 (s), 3.76 (d, *J* = 5.9 Hz), 7.30 (d, *J* = 6.1 Hz). LCMS m/z ES^+ $[M\text{-}Boc+H]^+$ 133.9.

3-Bromoprop-2-yn-1-amine hydrochloride



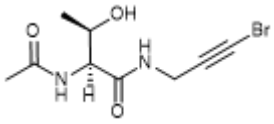
tert-Butyl (3-bromoprop-2-yn-1-yl)carbamate (1.1 g, 4.7 mmol) was dissolved in 4M HCl in dioxane (30 mL). The reaction mixture was stirred at room temperature for 2 h then evaporated to dryness to afford 3-bromoprop-2-yn-1-amine hydrochloride (0.79 g, 99%) as a yellow solid. m.p. 169 °C; IR ν_{max} (cm^{-1}) 2856, 2629, 2226, 2121, 2074; ^{1}H NMR (500 MHz, DMSO-d_6) δ 3.78 (s, 2H), 8.48 (s, 3H); LCMS m/z ES^+ $[M+H]^+$ 171.9.

(2*S*,3*R*)-2-Acetamido-*N*-(3-bromoprop-2-yn-1-yl)-3-(*tert*-butoxy)butanamide



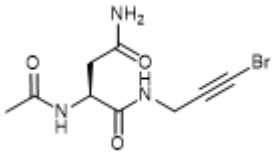
(2S,3S)-2-Acetamido-N-(3-bromoprop-2-yn-1-yl)-3-(tert-butoxy)butanamide was synthesized according to General procedure A using (2*S*,3*R*)-2-acetamido-3-(*tert*-butoxu)butanoic acid (0.41 g, 1.9 mmol). The crude product was purified by flash silica chromatography, elution gradient 0 – 10% MeOH in DCM. Pure fractions were evaporated to dryness to afford (2*S*,3*S*)-2-acetamido-*N*-(3-bromoprop-2-yn-1-yl)-3-(*tert*-butoxy)butanamide (0.20 g, 42%) as a white solid. R_f = 0.46 (10% MeOH in DCM); mp: 180 – 183 °C; IR ν_{max} (cm^{-1}) 3271, 3078, 2969, 2935, 2222, 2113; ^{1}H NMR (500 MHz, Methanol-d_4) δ 1.16 (d, *J* = 6.2, 5.0 Hz), 1.21 (s, *J* = 3.9 Hz, 9H), 2.01 (s, 3H), 3.91 – 4.09 (m, 3H), 4.32 (d, *J* = 7.5 Hz, 1H); ^{13}C NMR (126 MHz, Methanol-d_4) δ 18.61, 21.15, 27.27, 28.90, 41.92, 58.81, 67.21, 74.16, 75.57, 171.19, 171.92; LCMS m/z ES+ [M+H]+ 333.2; calcd for $C_{13}H_{21}{}^{79}BrN_2O_3$ 333.2260 $[M(Br)+H]^+$ found 333.0808.

(2*S*,3*R*)-2-Acetamido-*N*-(3-bromoprop-2-yn-1-yl)-3-hydroxybutanamide (threonine PepLite)



(2*S*,3*S*)-2-Acetamido-*N*-(3-bromoprop-2-yn-1-yl)-3-(*tert*-butoxy)butanamide (80 mg, 0.24 mmol) was dissolved in anhydrous DCM (20 mL) and TFA (10 mL) and 0 °C under nitrogen. The reaction mixture was allowed to warm to room temperature and was stirred at room temperature for 3 h then evaporated to dryness to afford crude product. The crude product was purified by flash silica chromatography, elution gradient 0 – 15% MeOH in DCM. Pure fractions were evaporated to dryness to afford (2*S*,3*S*)-2-acetamido-*N*-(3-bromoprop-2-yn-1-yl)-3-hydroxybutanamide (38 mg, 57%, 93% de) as a white solid. R_f = 0.34 (10% MeOH in DCM); mp: 189 – 192 °C; IR ν_{max} (cm^{-1}) 3280, 3085, 2973, 2924, 2225, 2115; ^{1}H NMR (500 MHz, Methanol-d_4) δ 1.21 (d, *J* = 6.4 Hz, 3H), 2.03 (s, 3H), 3.97 – 4.06 (m, 3H), 4.33 (d, *J* = 6.5 Hz, 1H); ^{13}C NMR (126 MHz, Methanol-d_4) δ 18.21, 21.13, 29.00, 41.79, 58.69, 67.11, 75.41, 170.88, 172.00; LCMS m/z ES^+ $[M+H]^+$ 277.1; calcd for $C_9H_{13}{}^{79}BrN_2O_3$ 277.1180 $[M(Br)+H]^+$ found 277.0182.

(*S*)-2-Acetamido-*N*^{1}-(3-bromoprop-2-yn-1-yl)succinimide (asparagine PepLite)



(*S*)-2-Acetamido-*N*^{1}-(3-bromoprop-2-yn-1-yl)succinamide was synthesized according to General procedure A using (*s*)-2-acetamido-5-amino-5-oxobutanoic acid (155 mg, 0.89 mmol) and evaporating the reaction mixture to afford the crude product without aqueous work-up. The crude product was purified by flash silica chromatography, elution gradients 0 – 10% MeOH in DCM. Pure fractions were evaporated to dryness to afford (*S*)-2-acetamido-*N*^{1}-(3-bromoprop-2-yn-1-yl)succinamide (50 mg, 30%) as a white solid. R_f = 0.18 (10% MeOH in DCM); mp: 173 °C (decomp); IR ν_{max} (cm^{-1}) 3421, 3277, 3208, 3072, 2922, 2226, 2116; 1H NMR

(500 MHz, Methanol-d4) δ 1.99 (s, 3H), 2.58 – 2.75 (m, 2H), 3.98 (d, J = 1.4 Hz, 2H), 4.71 (dd, J = 7.6, 5.7 Hz, 1H); 13C NMR (126 MHz, Methanol-d4) δ 22.57, 30.61, 37.83, 43.13, 51.54, 76.84, 173.04, 173.28, 174.81; LCMS m/z ES^+ $[M+H]^+$ 290.2.

Acknowledgements

We would like to thank all the staff of Diamond Light Source for providing support and encouragement which allowed us to carry out this work during the COVID-19 lockdown. We'd also like to thank the Diamond MX group for their support and expertise, in particular David Aragão, Ralf Flaig, Dave Hall, Katherine McAuley and Mark Williams. We are grateful to AstraZeneca, Astex Pharmaceuticals, Lilly, Pfizer and Vernalis, University of York (TDD, SPJ) and the EU (Horizon 2020 program, Marie Skłodowska-Curie grant agreement No. 675899, FRAGNET) (AK and HFK) for support. The SGC is a registered charity (number 1097737) that receives funds from AbbVie, Bayer Pharma AG, Boehringer Ingelheim, Canada Foundation for Innovation, Eshelman Institute for Innovation, Genome Canada, Innovative Medicines Initiative (EU/EFPIA) [ULTRA-DD grant no. 115766], Janssen, Merck KGaA Darmstadt Germany, MSD, Novartis Pharma AG, Ontario Ministry of Economic Development and Innovation, Pfizer, São Paulo Research Foundation-FAPESP, Takeda, and Wellcome [106169/ZZ14/Z]. N.L. is the incumbent of the Alan and Laraine Fischer Career Development Chair. N.L. would like to acknowledge funding from the Israel Science Foundation (grant no. 2462/19), The Israel Cancer Research Fund, the Israeli Ministry of Science Technology (grant no. 3-14763), and the Moross Integrated Cancer Center. N.L. is also supported by the Helen and Martin Kimmel Center for Molecular Design, Joel and Mady Dukler Fund for Cancer Research, the Estate of Emile Mimran and Virgin JustGiving, and the George Schwartzman Fund.

References

ANAND, K., ZIEBUHR, J., WADHWANI, P., MESTERS, J. R. & HILGENFELD, R. 2003. Coronavirus main proteinase (3CL(pro)) structure: Basis for design of anti-SARS drugs. *Science,* 300**,** 1763-1767.

ARROWSMITH, C. H., AUDIA, J. E., AUSTIN, C., BAELL, J., BENNETT, J., BLAGG, J., BOUNTRA, C., BRENNAN, P. E., BROWN, P. J., BUNNAGE, M. E., BUSER-DOEPNER, C., CAMPBELL, R. M., CARTER, A. J., COHEN, P., COPELAND, R. A., CRAVATT, B., DAHLIN, J. L., DHANAK, D., EDWARDS, A. M., FREDERIKSEN, M., FRYE, S. V., GRAY, N., GRIMSHAW, C. E., HEPWORTH, D., HOWE, T., HUBER, K. V., JIN, J., KNAPP, S., KOTZ, J. D., KRUGER, R. G., LOWE, D., MADER, M. M., MARSDEN, B., MUELLER-FAHRNOW, A., MULLER, S., O'HAGAN, R. C., OVERINGTON, J. P., OWEN, D. R., ROSENBERG, S. H., ROTH, B., ROSS, R., SCHAPIRA, M., SCHREIBER, S. L., SHOICHET, B., SUNDSTROM, M., SUPERTI-FURGA, G., TAUNTON, J., TOLEDO-SHERMAN, L., WALPOLE, C., WALTERS, M. A., WILLSON, T. M., WORKMAN, P., YOUNG, R. N. & ZUERCHER, W. J. 2015. The promise and peril of chemical probes. *Nat Chem Biol,* 11**,** 536-41.

BAELL, J. B. & HOLLOWAY, G. A. 2010. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem,* 53**,** 2719-40.

563 BAELL, J. B. & NISSINK, J. W. M. 2018. Seven Year Itch: Pan-Assay Interference Compounds
564 (PAINS) in 2017-Utility and Limitations. *ACS Chem Biol,* 13**,** 36-44.

565 BAILLIE, T. A. 2016. Targeted Covalent Inhibitors for Drug Design. *Angew Chem Int Ed Engl,*
566 55**,** 13408-13421.

567 BAJUSZ, D. & KESERÜ, G. M. *manuscript in preparation*.

568 BARRILA, J., BACHA, U. & FREIRE, E. 2006. Long-range cooperative interactions modulate
569 dimerization in SARS 3CL(pro). *Biochemistry,* 45**,** 14908-14916.

570 BAUER, R. A. 2015. Covalent inhibitors in drug discovery: from accidental discoveries to
571 avoided liabilities and designed therapies. *Drug Discov Today,* 20**,** 1061-73.

572 BAUMAN, J. D., HARRISON, J. J. E. K. & ARNOLD, E. 2016. Rapid experimental SAD phasing and
573 hot-spot identification with halogenated fragments. *Iucrj,* 3**,** 51-60.

574 BERMINGHAM, A., CHAND, M. A., BROWN, C. S., AARONS, E., TONG, C., LANGRISH, C.,
575 HOSCHLER, K., BROWN, K., GALIANO, M., MYERS, R., PEBODY, R. G., GREEN, H. K.,
576 BODDINGTON, N. L., GOPAL, R., PRICE, N., NEWSHOLME, W., DROSTEN, C., FOUCHIER,
577 R. A. & ZAMBON, M. 2012. Severe respiratory illness caused by a novel coronavirus, in
578 a patient transferred to the United Kingdom from the Middle East, September 2012.
579 *Eurosurveillance,* 17**,** 6-10.

580 BREDENBEEK, P. J., PACHUK, C. J., NOTEN, A. F. H., CHARITE, J., LUYTJES, W., WEISS, S. R. &
581 SPAAN, W. J. M. 1990. The Primary Structure and Expression of the 2nd Open Reading
582 Frame of the Polymerase Gene of the Coronavirus Mhv-A59 - a Highly Conserved
583 Polymerase Is Expressed by an Efficient Ribosomal Frameshifting Mechanism. *Nucleic*
584 *Acids Research,* 18**,** 1825-1832.

585 BRICOGNE, G., BLANC, E., BRANDL, M., FLENSBURG, C., KELLER, P., PACIOREK, W., ROVERSI,
586 P., SHARFF, A., SMART, O. S., VONRHEIN, C. & WOMACK, T. O. 2017. Buster. 2.10.13
587 ed. Cambridge, United Kingdom.

588 CHEN, S., ZHANG, J., HU, T. C., CHEN, K. X., JIANG, H. L. & SHEN, X. 2008. Residues on the
589 dimer interface of SARS coronavirus 3C-like protease: Dimer stability characterization
590 and enzyme catalytic activity analysis. *Journal of Biochemistry,* 143**,** 525-536.

591 CHOU, C. Y., CHANG, H. C., HSU, W. C., LIN, T. Z., LIN, C. H. & CHANG, G. G. 2004. Quaternary
592 structure of the severe acute respiratory syndrome (SARS) coronavirus main protease.
593 *Biochemistry,* 43**,** 14958-14970.

594 COLLINS, P. M., NG, J. T., TALON, R., NEKROSIUTE, K., KROJER, T., DOUANGAMATH, A.,
595 BRANDAO-NETO, J., WRIGHT, N., PEARCE, N. M. & VON DELFT, F. 2017. Gentle, fast
596 and effective crystal soaking by acoustic dispensing. *Acta Crystallographica Section D-*
597 *Structural Biology,* 73**,** 246-255.

598 COX, O. B., KROJER, T., COLLINS, P., MONTEIRO, O., TALON, R., BRADLEY, A., FEDOROV, O.,
599 AMIN, J., MARSDEN, B. D., SPENCER, J., VON DELFT, F. & BRENNAN, P. E. 2016. A
600 poised fragment library enables rapid synthetic expansion yielding the first reported
601 inhibitors of PHIP(2), an atypical bromodomain. *Chem Sci,* 7**,** 2322-2330.

602 DE CESCO, S., KURIAN, J., DUFRESNE, C., MITTERMAIER, A. K. & MOITESSIER, N. 2017. Covalent
603 inhibitors design and discovery. *Eur J Med Chem,* 138**,** 96-114.

604 DONG, E. S., DU, H. R. & GARDNER, L. 2020. An interactive web-based dashboard to track
605 COVID-19 in real time. *Lancet Infectious Diseases,* 20**,** 533-534.

606 DOWNES, T. D., JONES, S. P., KLEIN, H. F., WHELDON, M. C., ATOBE, M., BOND, P. S., FIRTH, J.
607 D., CHAN, N. S., WADDELOVE, L., HUBBARD, R. E., BLAKEMORE, D. C., DE FUSCO, C.,
608 ROUGHLEY, S. D., VIDLER, L. R., WHATTON, M. A., WOOLFORD, A. J., WRIGLEY, G. L. &
609 O'BRIEN, P. 2020. Design and Synthesis of 56 Shape Diverse 3-D Fragments. *Chem. Eur.*
610 *J.,* published online ahead of print, 2020 Apr 21.

611 EMSLEY, P., LOHKAMP, B., SCOTT, W. G. & COWTAN, K. 2010. Features and development of
612 Coot. *Acta Crystallographica Section D-Biological Crystallography,* 66**,** 486-501.

613 EVANS, P. R. & MURSHUDOV, G. N. 2013. How good are my data and what is the resolution?
614 *Acta Crystallographica Section D-Biological Crystallography,* 69**,** 1204-1214.

615 FERLA, M. P., PAGNAMENTA, A. T., DAMERELL, D., TAYLOR, J. C. & MARSDEN, B. D. 2020.
616 MichelaNglo: sculpting protein views on web pages without coding. *Bioinformatics,*
617 36**,** 3268-3270.

618 FULLER, N., SPADOLA, L., COWEN, S., PATEL, J., SCHONHERR, H., CAO, Q., MCKENZIE, A.,
619 EDFELDT, F., RABOW, A. & GOODNOW, R. 2016. An improved model for fragment-
620 based lead generation at AstraZeneca. *Drug Discov Today,* 21**,** 1272-83.

621 GHOSH, A. K., XI, K., GRUM-TOKARS, V., XU, X. M., RATIA, K., FU, W. T., HOUSER, K. V., BAKER,
622 S. C., JOHNSON, M. E. & MESECAR, A. D. 2007. Structure-based design,synthesis, and
623 biological evaluation of peptidomimetic SARS-CoV 3CLpro inhibitors. *Bioorganic &*
624 *Medicinal Chemistry Letters,* 17**,** 5876-5880.

625 HEGYI, A. & ZIEBUHR, J. 2002. Conservation of substrate specificities among coronavirus main
626 proteases. *Journal of General Virology,* 83**,** 595-599.

627 HILGENFELD, R. 2014. From SARS to MERS: crystallographic studies on coronaviral proteases
628 enable antiviral drug design. *Febs Journal,* 281**,** 4085-4096.

629 HSU, W. C., CHANG, H. C., CHOU, C. Y., TSAI, P. J., LIN, P. I. & CHANG, G. G. 2005. Critical
630 assessment of important regions in the subunit association and catalytic action of the
631 severe acute respiratory syndrome coronavirus main protease. *Journal of Biological*
632 *Chemistry,* 280**,** 22741-22748.

633 HU, T. C., ZHANG, Y., LI, L. W., WANG, K. F., CHEN, S., CHEN, J., DING, J. P., JIANG, H. L. & SHEN,
634 X. 2009. Two adjacent mutations on the dimer interface of SARS coronavirus 3C-like
635 protease cause different conformational changes in crystal structure. *Virology,* 388**,**
636 324-334.

637 JIN, Z., DU, X., XU, Y., DENG, Y., LIU, M., ZHAO, Y., ZHANG, B., LI, X., ZHANG, L., PENG, C., DUAN,
638 Y., YU, J., WANG, L., YANG, K., LIU, F., JIANG, R., YANG, X., YOU, T., LIU, X., YANG, X.,
639 BAI, F., LIU, H., LIU, X., GUDDAT, L. W., XU, W., XIAO, G., QIN, C., SHI, Z., JIANG, H.,
640 RAO, Z. & YANG, H. 2020. Structure of M(pro) from SARS-CoV-2 and discovery of its
641 inhibitors. *Nature*.

642 KABSCH, W. 2010. Integration, scaling, space-group assignment and post-refinement. *Acta*
643 *Crystallographica Section D-Biological Crystallography,* 66**,** 133-144.

644 KEEGAN, R., WOJDYR, M., WINTER, G. & ASHTON, A. 2015. DIMPLE: A difference map pipeline
645 for the rapid screening of crystals on the beamline. *Acta Crystallographica a-*
646 *Foundation and Advances,* 71**,** S18-S18.

KEELEY, A., ABRANYI-BALOGH, P., HRAST, M., IMRE, T., ILAS, J., GOBEC, S. & KESERU, G. M. 2018. Heterocyclic electrophiles as new MurA inhibitors. *Arch Pharm (Weinheim),* 351**,** e1800184.

KEELEY, A., ABRANYI-BALOGH, P. & KESERU, G. M. 2019. Design and characterization of a heterocyclic electrophilic fragment library for the discovery of cysteine-targeted covalent inhibitors. *Medchemcomm,* 10**,** 263-267.

KROJER, T., TALON, R., PEARCE, N., COLLINS, P., DOUANGAMATH, A., BRANDAO-NETO, J., DIAS, A., MARSDEN, B. & VON DELFT, F. 2017. The XChemExplorer graphical workflow tool for routine or large-scale protein-ligand structure determination. *Acta Crystallographica Section D-Structural Biology,* 73**,** 267-278.

KUCHARSKI, A. J., RUSSELL, T. W. & DIAMOND, C. 2020. Early dynamics of transmission and control of COVID-19: a mathematical modelling study (vol 20, pg 553, 2020). *Lancet Infectious Diseases,* 20**,** E79-E79.

KUHN, B., GILBERG, E., TAYLOR, R., COLE, J. & KORB, O. 2019. How Significant Are Unusual Protein-Ligand Interactions? Insights from Database Mining. *Journal of Medicinal Chemistry,* 62**,** 10441-10455.

KUIKEN, T., FOUCHIER, R. A. M., SCHUTTEN, M., RIMMELZWAAN, G. F., VAN AMERONGEN, G., VAN RIEL, D., LAMAN, J. D., DE JONG, T., VAN DOORNUM, G., LIM, W., LING, A. E., CHAN, P. K. S., TAM, J. S., ZAMBON, M. C., GOPAL, R., DROSTEN, C., VAN DER WERF, S., ESCRIOU, N., MANUGUERRA, J. C., STOHR, K., PEIRIS, J. S. M. & OSTERHAUS, A. D. M. E. 2003. Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. *Lancet,* 362**,** 263-270.

LEE, W. H., YUE, W. W., RAUSH, E., TOTROV, M., ABAGYAN, R., OPPERMANN, U. & MARSDEN, B. D. 2011. Interactive JIMD articles using the iSee concept: turning a new page on structural biology data. *J Inherit Metab Dis,* 34**,** 565-7.

LONG, F., NICHOLLS, R. A., EMSLEY, P., GRAZULIS, S., MERKYS, A., VAITKUS, A. & MURSHUDOV, G. N. 2017. AceDRG: a stereochemical description generator for ligands. *Acta Crystallographica Section D-Structural Biology,* 73**,** 112-122.

MONS, E., JANSEN, I. D. C., LOBODA, J., VAN DOODEWAERD, B. R., HERMANS, J., VERDOES, M., VAN BOECKEL, C. A. A., VAN VEELEN, P. A., TURK, B., TURK, D. & OVAA, H. 2019. The Alkyne Moiety as a Latent Electrophile in Irreversible Covalent Small Molecule Inhibitors of Cathepsin K. *Journal of the American Chemical Society,* 141**,** 3507-3514.

MURSHUDOV, G. N., SKUBAK, P., LEBEDEV, A. A., PANNU, N. S., STEINER, R. A., NICHOLLS, R. A., WINN, M. D., LONG, F. & VAGIN, A. A. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D-Structural Biology,* 67**,** 355-367.

NOBLE, M. & WARING, M. *manuscript in preparation*.

O'REILLY, M., CLEASBY, A., DAVIES, T. G., HALL, R. J., LUDLOW, R. F., MURRAY, C. W., TISI, D. & JHOTI, H. 2019. Crystallographic screening using ultra-low-molecular-weight ligands to guide drug design. *Drug Discov Today,* 24**,** 1081-1086.

PEARCE, N. M., KROJER, T., BRADLEY, A. R., COLLINS, P., NOWAK, R. P., TALON, R., MARSDEN, B. D., KELM, S., SHI, J. Y., DEANE, C. M. & VON DELFT, F. 2017. A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nature Communications,* 8.

691 RESNICK, E., BRADLEY, A., GAN, J. R., DOUANGAMATH, A., KROJER, T., SETHI, R., GEURINK, P.
692 P., AIMON, A., AMITAI, G., BELLINI, D., BENNETT, J., FAIRHEAD, M., FEDOROV, O.,
693 GABIZON, R., GAN, J., GU, J. X., PLOTNIKOV, A., REZNIK, N., RUDA, G. F., DIAZ-SAEZ, L.,
694 STRAUB, V. M., SZOMMER, T., VELUPILLAI, S., ZAIDMAN, D., ZHANG, Y. L., COKER, A.
695 R., DOWSON, C. G., BARR, H. M., WANG, C., HUBER, K. V. M., BRENNAN, P. E., OVAA,
696 H., VON DELFT, F. & LONDON, N. 2019. Rapid Covalent-Probe Discovery by
697 Electrophile-Fragment Screening. *Journal of the American Chemical Society,* 141**,**
698 8951-8968.

699 SHI, J. H. & SONG, J. X. 2006. The catalysis of the SARS 3C-like protease is under extensive
700 regulation by its extra domain. *Febs Journal,* 273**,** 1035-1045.

701 SINGH, J., PETTER, R. C., BAILLIE, T. A. & WHITTY, A. 2011. The resurgence of covalent drugs.
702 *Nat Rev Drug Discov,* 10**,** 307-17.

703 SIROIS, S., HATZAKIS, G., WEI, D., DU, Q. & CHOU, K. C. 2005. Assessment of chemical libraries
704 for their druggability. *Comput Biol Chem,* 29**,** 55-67.

705 SMART, O. S., WOMACK, T. O., SHARFF, A., FLENSBURG, C., KELLER, P., PACIOREK, W.,
706 VONRHEIN, C. & BRICOGNE, G. 2010. grade. 1.2.19 ed. Cambridge, United Kingdom:
707 Global Phasing Ltd.

708 STADLER, K., MASIGNANI, V., EICKMANN, M., BECKER, S., ABRIGNANI, S., KLENK, H. D. &
709 RAPPUOLI, R. 2003. SARS - Beginning to understand a new virus. *Nature Reviews*
710 *Microbiology,* 1**,** 209-218.

711 THIEL, V., IVANOV, K. A., PUTICS, A., HERTZIG, T., SCHELLE, B., BAYER, S., WEISSBRICH, B.,
712 SNIJDER, E. J., RABENAU, H., DOERR, H. W., GORBALENYA, A. E. & ZIEBUHR, J. 2003.
713 Mechanisms and enzymes involved in SARS coronavirus genome expression. *Journal*
714 *of General Virology,* 84**,** 2305-2315.

715 THOMAS, S. E., COLLINS, P., JAMES, R. H., MENDES, V., CHAROENSUTTHIVARAKUL, S.,
716 RADOUX, C., ABELL, C., COYNE, A. G., FLOTO, R. A., VON DELFT, F. & BLUNDELL, T. L.
717 2019. Structure-guided fragment-based drug discovery at the synchrotron: screening
718 binding sites and correlations with hotspot mapping. *Philosophical Transactions of the*
719 *Royal Society a-Mathematical Physical and Engineering Sciences,* 377.

720 VERSCHUEREN, K. H. G., PUMPOR, K., ANEMULLER, S., CHEN, S., MESTERS, J. R. &
721 HILGENFELD, R. 2008. A structural view of the inactivation of the SARS coronavirus
722 main proteinase by benzotriazole esters. *Chemistry & Biology,* 15**,** 597-606.

723 VONRHEIN, C., FLENSBURG, C., KELLER, P., SHARFF, A., SMART, O., PACIOREK, W., WOMACK,
724 T. & BRICOGNE, G. 2011. Data processing and analysis with the autoPROC toolbox.
725 *Acta Crystallographica Section D-Structural Biology,* 67**,** 293-302.

726 WINTER, G., LOBLEY, C. M. C. & PRINCE, S. M. 2013. Decision making in xia2. *Acta*
727 *Crystallographica Section D-Biological Crystallography,* 69**,** 1260-1273.

728 WINTER, G., WATERMAN, D. G., PARKHURST, J. M., BREWSTER, A. S., GILDEA, R. J., GERSTEL,
729 M., FUENTES-MONTERO, L., VOLLMAR, M., MICHELS-CLARK, T., YOUNG, I. D., SAUTER,
730 N. K. & EVANS, G. 2018. DIALS: implementation and evaluation of a new integration
731 package. *Acta Crystallographica Section D-Structural Biology,* 74**,** 85-97.

732 WOOD, D. J., LOPEZ-FERNANDEZ, J. D., KNIGHT, L. E., AL-KHAWALDEH, I., GAI, C. H., LIN, S. Y.,
733 MARTIN, M. P., MILLER, D. C., CANO, C., ENDICOTT, J. A., HARDCASTLE, I. R., NOBLE,
734 M. E. M. & WARING, M. J. 2019. FragLites-Minimal, Halogenated Fragments Displaying

735 Pharmacophore Doublets. An Efficient Approach to Druggability Assessment and Hit
736 Generation. *Journal of Medicinal Chemistry,* 62**,** 3741-3752.

737 WU, F., ZHAO, S., YU, B., CHEN, Y. M., WANG, W., SONG, Z. G., HU, Y., TAO, Z. W., TIAN, J. H.,
738 PEI, Y. Y., YUAN, M. L., ZHANG, Y. L., DAI, F. H., LIU, Y., WANG, Q. M., ZHENG, J. J., XU,
739 L., HOLMES, E. C. & ZHANG, Y. Z. 2020. A new coronavirus associated with human
740 respiratory disease in China. *Nature,* 579**,** 265-+.

741 XUE, X. Y., YU, H. W., YANG, H. T., XUE, F., WU, Z. X., SHEN, W., LI, J., ZHOU, Z., DING, Y., ZHAO,
742 Q., ZHANG, X. J. C., LIAO, M., BARTLAM, M. & RAO, Z. 2008. Structures of two
743 coronavirus main proteases: Implications for substrate binding and antiviral drug
744 design. *Journal of Virology,* 82**,** 2515-2527.

745 YANG, H. T., XIE, W. Q., XUE, X. Y., YANG, K. L., MA, J., LIANG, W. X., ZHAO, Q., ZHOU, Z., PEI,
746 D. Q., ZIEBUHR, J., HILGENFELD, R., YUEN, K. Y., WONG, L., GAO, G. X., CHEN, S. J.,
747 CHEN, Z., MA, D. W., BARTLAM, M. & RAO, Z. 2005. Design of wide-spectrum inhibitors
748 targeting coronavirus main proteases. *Plos Biology,* 3**,** 1742-1752.

749 YANG, H. T., YANG, M. J., DING, Y., LIU, Y. W., LOU, Z. Y., ZHOU, Z., SUN, L., MO, L. J., YE, S.,
750 PANG, H., GAO, G. F., ANAND, K., BARTLAM, M., HILGENFELD, R. & RAO, Z. H. 2003.
751 The crystal structures of severe acute respiratory syndrome virus main protease and
752 its complex with an inhibitor. *Proceedings of the National Academy of Sciences of the*
753 *United States of America,* 100**,** 13190-13195.

754 ZAKI, A. M., VAN BOHEEMEN, S., BESTEBROER, T. M., OSTERHAUS, A. D. M. E. & FOUCHIER, R.
755 A. M. 2012. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi
756 Arabia. *New England Journal of Medicine,* 367**,** 1814-1820.

757 ZHANG, L., LIN, D., KUSOV, Y., NIAN, Y., MA, Q., WANG, J., VON BRUNN, A., LEYSSEN, P.,
758 LANKO, K., NEYTS, J., DE WILDE, A., SNIJDER, E. J., LIU, H. & HILGENFELD, R. 2020a.
759 alpha-Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus
760 Replication: Structure-Based Design, Synthesis, and Activity Assessment. *J Med Chem,*
761 63**,** 4562-4578.

762 ZHANG, L. L., LIN, D. Z., SUN, X. Y. Y., CURTH, U., DROSTEN, C., SAUERHERING, L., BECKER, S.,
763 ROX, K. & HILGENFELD, R. 2020b. Crystal structure of SARS-CoV-2 main protease
764 provides a basis for design of improved alpha-ketoamide inhibitors. *Science,* 368**,** 409-
765 +.

766 ZHANG, T., HATCHER, J. M., TENG, M., GRAY, N. S. & KOSTIC, M. 2019. Recent Advances in
767 Selective and Irreversible Covalent Ligand Development and Validation. *Cell Chem*
768 *Biol,* 26**,** 1486-1500.

769 ZHU, N., ZHANG, D. Y., WANG, W. L., LI, X. W., YANG, B., SONG, J. D., ZHAO, X., HUANG, B. Y.,
770 SHI, W. F., LU, R. J., NIU, P. H., ZHAN, F. X., MA, X. J., WANG, D. Y., XU, W. B., WU, G.
771 Z., GAO, G. G. F., TAN, W. J. & CORONAVIRUS, C. N. 2020. A Novel Coronavirus from
772 Patients with Pneumonia in China, 2019. *New England Journal of Medicine,* 382**,** 727-
773 733.

774