



STRUCTURAL
BIOLOGY

Volume 80 (2024)

Supporting information for article:

Fragment-based screening targeting an open form of the SARS-CoV-2 main protease binding pocket

Chia-Ying Huang, Alexander Metz, Roland Lange, Nadia Artico, Céline Potot, Julien Hazemann, Manon Müller, Marina Dos Santos, Alain Chambovey, Daniel Ritz, Deniz Eris, Solange Meyer, Geoffroy Bourquin, May Sharpe and Aengus Mac Sweeney

S1. Cloning and expression

SARS-CoV-2 3CL^{pro} expression as a Hexa-histidine-SUMO-fusion protein
amino acid sequence:

MGSSHHHHHGSGLVPRGSASMSDSEVDQEAKPEVKPEVKPETHINLKVSDGSS
EIFFKIKKTTPLRRLMEAFQKRQKEMDSLRFYDGIQADQTPEDLDMEDND
IEAHREQIGGSGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCPRHV
ICTSEDMLNPYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVLKLKVD TANP
KTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSCGSVGF
NIDYDCVSFCYMHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVN
VLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNIEPLTQDHVDILGPLSAQ
TGI AVLDMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGVTFQ

Black: Hexa-His tag and linker ; Blue: SUMO ; Red: 3CL^{pro}

S2. Hexahistidine SUMO-3CL^{pro}

425 aas; Mol Wt 47194.9, Isoelectric Pt (pI) 5.80

Amino acid sequence “Backtranseq”; *E. coli* codon usage high

https://www.ebi.ac.uk/Tools/st/emboss_backtranseq/

S3. DNA synthesized (Genscript) and cloned into pET29a+

CATATGGGTTCTTCTCACCACCACCACCACCGTTCTGGTCTGGTTCCGCGTG
GTTCTGCGTCTATGTCTGACTCTGAAGTTGACCAGGAAGCGAAACCGGAAGTTAA
ACCGGAAGTTAAACCGGAACCCACATCAACCTGAAAGTTTCTGACGGTTCTTCT
GAAATCTTCTTCAAAATCAAAAAAACCACCCCGCTGCGTCGTCTGATGGAAGCGT
TCGCGAAACGTCAGGGTAAAGAAATGGACTCTCTGCGTTTCCTGTACGACGGTAT
CCGTATCCAGGCGGACCAGACCCCGGAAGACCTGGACATGGAAGACAACGACAT
CATCGAAGCGCACCGTGAACAGATCGGTGGTTCTGGTTTCCGTAAAATGGCGTTC
CCGTCTGGTAAAGTTGAAGGTTGCATGGTTCAGGTTACCTGCGGTACCACCACCC

TGAACGGTCTGTGGCTGGACGACGTTGTTTACTGCCCCGCGTCACGTTATCTGCAC
CTCTGAAGACATGCTGAACCCGAACCTACGAAGACCTGCTGATCCGTAAATCTAAC
CACAACCTCCTGGTTCAGGCGGGTAACGTTTACGCTGCGTGTTATCGGTCACTCTA
TGCAGAACTGCGTTCTGAAACTGAAAGTTGACACCGCGAACCCGAAAACCCCGA
AATACAAATTCGTTTCGTATCCAGCCGGGTCAGACCTTCTCTGTTCTGGCGTGCTA
CAACGGTTCTCCGTCTGGTGTTTACCAGTGCGCGATGCGTCCGAACCTTCACCATC
AAAGGTTCTTTCCTGAACGGTTCTTGCGGTTCTGTTGGTTTCAACATCGACTACGA
CTGCGTTTCTTTCTGCTACATGCACCACATGGAACCTGCCGACCGGTGTTACGCG
GGTACCGACCTGGAAGGTAACTTCTACGGTCCGTTTCGTTGACCGTCAGACCGCGC
AGGCGGCGGGTACCGACACCACCATCACCGTTAACGTTCTGGCGTGGCTGTACGC
GGCGGTTATCAACGGTGACCGTTGGTTCCTGAACCGTTTACCACCACCCTGAAC
GACTTCAACCTGGTTGCGATGAAATACAACCTACGAACCGCTGACCCAGGACCAC
GTTGACATCCTGGGTCCGCTGTCTGCGCAGACCGGTATCGCGGTTCTGGACATGT
GCGCGTCTCTGAAAGAACTGCTGCAGAACGGTATGAACGGTCGTACCATCCTGG
GTTCTGCGCTGCTGGAAGACGAATTCACCCCGTTCGACGTTGTTTCGTCAGTGCTC
TGGTGTTACCTTCCAGTTAATAGGGATCC

S4. Storage buffer (X-ray)

20 mM Tris-HCl pH 7.8, 150 mM NaCl, 1 mM TCEP, 1 mM EDTA

S5. Storage buffer (Rapidfire MS)

20 mM Tris-HCl pH 7.8, 150 mM NaCl, 1 mM TCEP, 1 mM EDTA, 5% glycerol

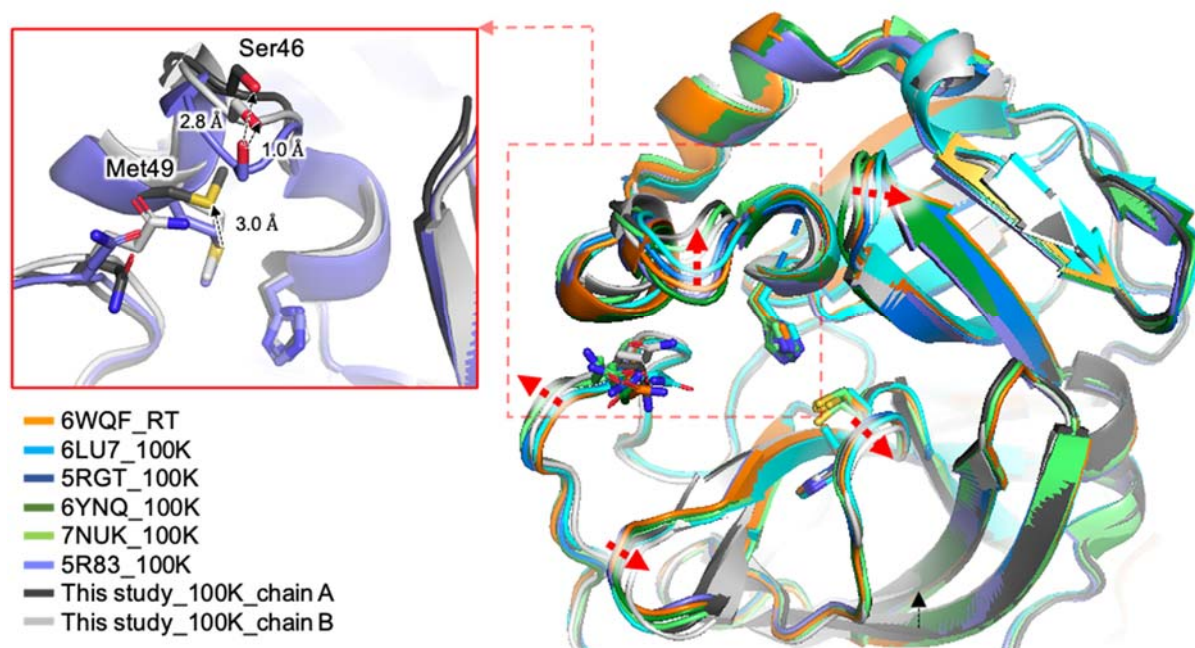


Figure S1 Comparison of the active site cavity conformations of different 3CL^{pro} structures. The protein structures are shown in cartoon representation, with selected active site residues shown in a stick representation. The distance is indicated by dashed arrows and corresponding residues are highlighted.

6WQF: Kneller et al 2020 (<https://doi.org/10.1038/s41467-020-16954-7>)

6LU7: Jin et al 2020 (<https://doi.org/10.1038/s41586-020-2223-y>)

5RGT: Zaidman et al 2021 (<https://doi.org/10.1016/j.chembiol.2021.05.018>)

6YNQ: Gunther et al., 2021 (<https://doi.org/10.1126/science.abf7945>)

7NUK: Sutanto et al., 2021 (<https://doi.org/10.1002/anie.202105584>)

5R83: Douangamath et al., 2020 (<https://doi.org/10.1038/s41467-020-18709-w>)

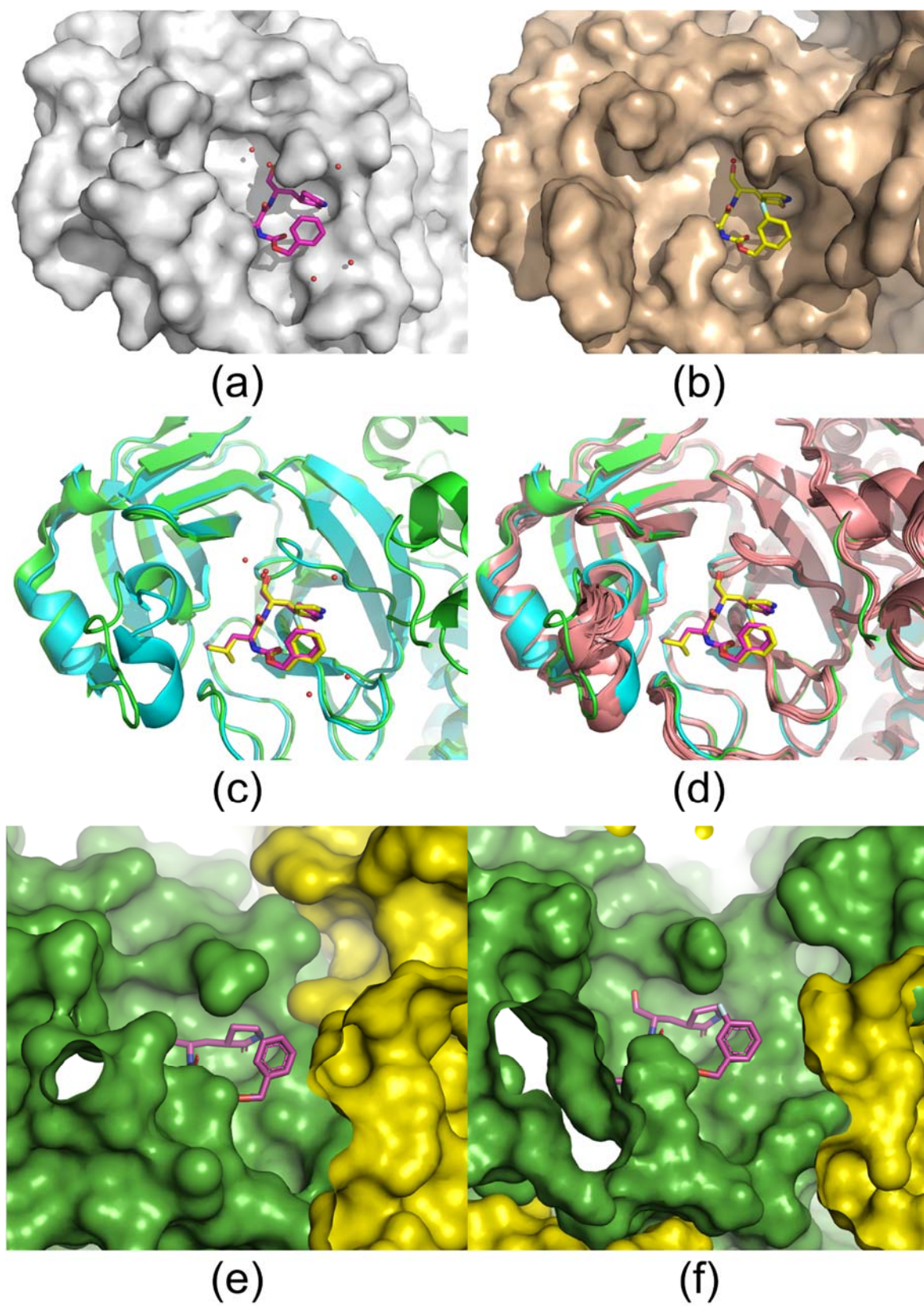


Figure S2 Pocket openness and accessibility. Depicted are (a) the pocket conformations of PDB codes 7c6u (“Type 1” crystals) and (b) 7lcr (“Type 2” crystals) with the protein in surface representation and the ligands as stick models. In (c) both structures are superimposed as cartoon models highlighting the similarly open pocket conformation, except for the left side rim (Cys44 to Ans51, see arrow) that is disordered in 7lcr (green) compared to the helix found for this stretch in 7c6u (cyan). In (d), the additional superimposition with all fragment-bound structures reported in the study (rose cartoon) reveals similar pocket conformations with the majority exhibiting an ordered rim section. Also the crystal packing close to the pocket, and hence the accessibility, is similar for both crystal forms as shown for (e) 7c6u and for (f) 7lcr, with surface representations for the asymmetric unit containing the pocket of interest (green) and close-by symmetry mates (yellow).

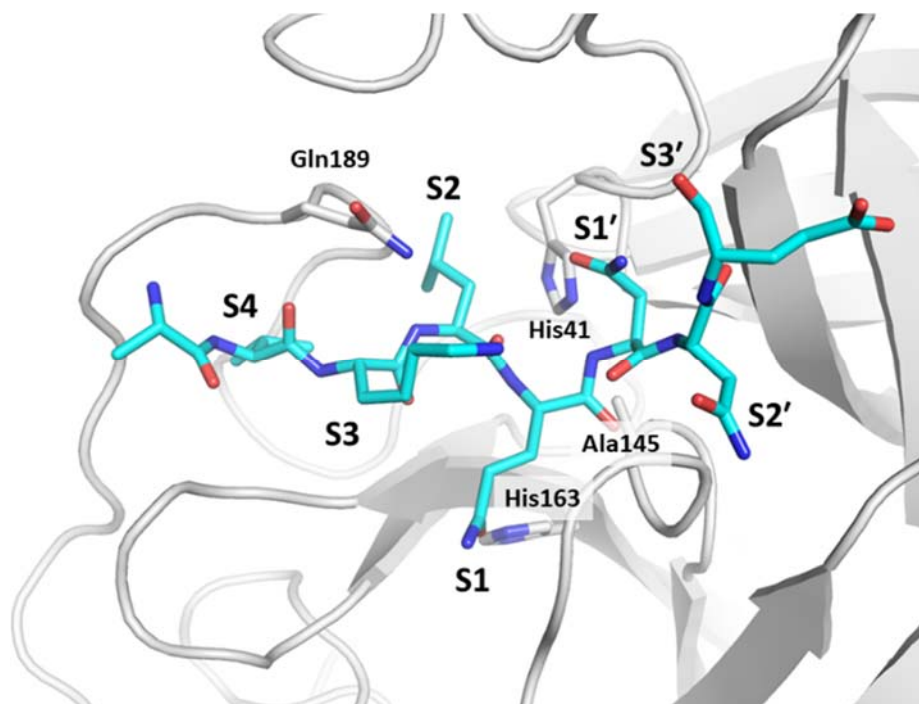


Figure S3 A substrate peptide bound to a catalytically inactive mutant of 3CL^{pro} (PDB entry 7mgr). The protein is shown in cartoon representation in grey, and the substrate peptide is shown in stick representations in blue. Stick models show C (cyan), N (blue) and O (red) atoms.

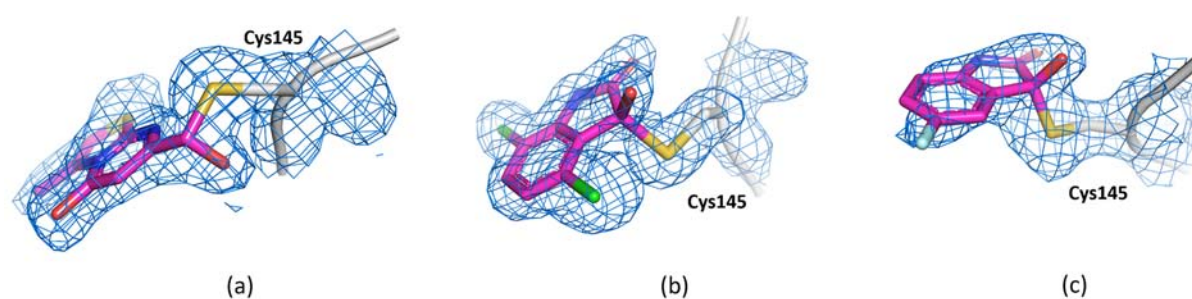


Figure S4 2Fo-Fc electron density contoured at one sigma, showing the covalent bond of (a) cpd-27 (b) cpd-28 and (c) cpd-29 to Cys145 of 3CL^{pro}. The protein is shown in cartoon representation in grey, with the Cys145 side chain in stick representation, and the inhibitors are shown in stick representations in magenta. Stick models show C (magenta), N (blue), O (red), F (light blue) and Cl (green) atoms.

Table S1. Data collection and refinement statistics.

Data collection*	cpd-1	cpd-2	cpd-3	cpd-4	cpd-5	cpd-6	cpd-7	cpd-8	cpd-9	cpd-10	cpd-11	cpd-12	cpd-13	cpd-14	cpd-15	cpd-16	cpd-17	cpd-18	cpd-19	cpd-20	cpd-21	cpd-22	cpd-23	cpd-24	cpd-25	cpd-26	cpd-27	cpd-28	cpd-29
PTD code	70RE	70RF	70RG	70RH	70RI	70RJ	70RK	70RL	70RM	70RN	70RO	70RP	70RQ	70RR	70RS	70RT	70RU	70RV	70RW	70RX	70RY	70RZ	70RO	70RI	70R2	70R3	70R4	70R5	70R6
Resolution range (Å)	56.82- 1.66	56.82- 1.84	56.68- 1.54	56.94- 1.87	57.09- 1.79	46.36- 1.74	56.25- 1.80	56.77- 1.68	56.65- 1.70	56.46- 1.92	56.68- 1.55	56.70- 1.56	56.70- 1.67	52.04- 1.68	45.93- 1.47	56.07- 1.71	56.71- 1.31	56.82- 1.92	56.94- 1.92	56.08- 1.55	56.64- 1.54	56.82- 1.86	55.81- 1.69	71.51- 1.74	56.21- 1.77	56.85- 1.89	56.85- 1.88	56.94- 1.88	56.16- 1.62
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions a, b, c (Å)	67.79 99.93 104.17	67.97 99.47 104.18	67.78 99.47 104.17	67.97 99.47 104.17	68.03 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18	67.83 99.47 104.18
Total reflections	850712	851538	1169474	579937	712447	672153	712405	904775	904904	536788	988735	929045	801741	799019	1120370	705566	780617	684097	5530002	1131064	1004192	837150	923113	669878	821338	440297	622297	556557	1066123
Unique reflections	(39169)	(26315)	(57088)	(36987)	(32279)	(34622)	(40174)	(44166)	(44166)	(30350)	(43164)	(67469)	(34103)	(37171)	(34073)	(30303)	(40678)	(35913)	(27880)	(55598)	(43267)	(134943)	(44141)	(28593)	(41410)	(20920)	(32253)	(23282)	(51011)
Completeness spherical (%)***	99.7	99.6	99.9	99.7	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8
Completeness ellipsoidal (%)	99.7	99.6	99.9	99.7	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8
Multiplicity	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6
Refinement	62370	42759	85946	42624	51554	50966	52012	68505	66548	40980	70608	67641	60576	58285	83797	52109	56541	49735	40997	84280	73816	58522	67558	49437	59678	34289	46076	40565	77251
Number of non-hydrogen atoms	5324	5142	5313	5063	5110	5091	5176	5258	5118	5074	5394	5362	5241	4970	5574	5006	5242	5132	5101	5387	5336	5112	5208	5193	5276	5088	5122	4945	5275
R _{int} / R _{obs}	0.25	0.24	0.24	0.25	0.26	0.27	0.24	0.23	0.23	0.23	0.23	0.23	0.23	0.25	0.22	0.26	0.14	0.24	0.24	0.22	0.22	0.25	0.22	0.25	0.24	0.28	0.28	0.25	0.23
macromolecules	4722	4687	4730	4694	4664	4691	4670	4721	4686	4658	4777	4724	4693	4706	4734	4682	4718	4697	4654	4739	4749	4663	4732	4691	4705	4709	4684	4676	4707
ligands	58	54	36	89	37	49	35	46	66	59	76	45	45	45	129	56	52	57	70	82	54	64	41	54	52	35	36	83	69
solvent	544	401	547	280	409	351	471	491	366	357	541	593	501	219	711	268	472	378	377	566	533	385	435	448	519	344	402	186	499
RMS(bonds) (Å)	0.008	0.004	0.007	0.006	0.004	0.004	0.008	0.019	0.020	0.009	0.009	0.022	0.010	0.011	0.009	0.019	0.097	0.012	0.005	0.008	0.015	0.005	0.011	0.066	0.010	0.003	0.004	0.010	0.013
RMS(angles) (°)	0.96	0.67	0.98	0.86	0.61	0.66	1.03	1.22	1.04	1.08	1.08	1.79	1.17	1.17	1.06	0.89	3.37	1.29	0.79	0.97	1.43	0.77	1.22	2.71	1.09	0.52	1.02	1.10	1.40
Ramachandran favored (%)	97.86	97.51	98.19	97.35	98.16	97.84	97.68	97.68	98.16	97.16	98.52	98.51	98.50	97.82	98.03	98.00	98.51	97.84	98.16	97.84	98.01	97.83	98.84	98.84	98.02	96.86	98.17	96.01	98.17
Ramachandran allowed (%)	2.14	2.49	1.81	2.65	1.84	2.16	2.32	1.84	1.48	1.49	1.50	1.50	1.50	2.18	1.97	2.00	1.49	2.16	1.84	2.16	1.99	2.17	1.32	1.16	1.98	3.14	1.83	3.99	1.83
Ramachandran outliers (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Clashscore	4.02	4.26	3.72	5.59	2.90	4.49	4.41	3.71	3.19	3.95	4.02	5.23	5.65	4.88	4.06	5.82	6.49	4.93	2.83	3.16	3.16	3.33	4.13	3.30	5.00	4.80	6.11	6.71	4.78
Average B-factor	22.00	28.20	25.40	35.80	29.04	34.09	27.97	24.44	28.44	30.21	25.70	24.48	20.81	28.15	23.67	27.15	24.89	32.21	36.26	25.97	24.44	38.24	28.17	25.52	21.72	19.49	33.75	38.80	28.18
macromolecules	21.12	27.72	24.52	35.25	28.55	33.95	27.31	23.55	27.55	30.28	24.56	23.32	19.95	27.94	21.81	26.63	24.02	31.70	35.64	24.62	24.17	37.70	27.33	24.80	20.78	19.42	33.42	38.50	27.02
ligands	33.19	41.67	34.50	62.79	39.31	42.38	35.02	35.65	57.47	34.54	46.07	40.07	31.04	43.57	44.11	49.37	46.19	43.58	51.36	40.14	45.30	46.57	33.39	38.21	32.81	27.24	44.27	59.70	49.64
solvent	28.41	32.01	32.43	36.40	33.70	34.73	33.97	31.92	34.58	28.50	32.92	32.52	27.92	29.38	32.30	31.45	31.84	36.80	41.05	35.22	34.79	43.30	36.82	31.52	29.19	19.72	36.66	36.93	36.18

*Data processing statistics are reported with Friedel pairs separated.

**Statistics for the highest-resolution shell are shown in parentheses.

***With the exception of the cpd-22, which was processed by XDS, datasets were processed and elliptically truncated using the autoPROC STANISISO option.

Table S1 SMILES codes, occupancy and 2mFo-DFc real space correlation coefficient (RSCC, from PDB validation report) of 3CL^{pro} fragment hits. Where multiple copies of the ligand are present in the structure, the highest RSCC value is shown.

* Fragments bound to the protein in different locations.

** Fragments bound to the protein in different conformations at the same location.

Compound	SMILES code	Occupancy	RSCC
1	<chem>FC(c1cc(-c2ccncc2)n[nH]1)(F)F</chem>	0.83	0.76
2	<chem>Nc1cc(Br)cnc1</chem>	0.70/0.61*	0.68
3	<chem>Nc(c(Cl)cnc1)c1Cl</chem>	0.77/0.67*	0.76
4	<chem>Oc1cc(Cl)cnc1</chem>	0.81/0.73*	0.88
5	<chem>CC(c1ccn[nH]1)=O.Cl</chem>	0.84/0.71*	0.89
6	<chem>OCc1c[s]c(cc2)c1cc2Cl</chem>	0.72/0.82*	0.92
7	<chem>OCc1cc(F)cc2c1OCOC2</chem>	0.80	0.88
8	<chem>CC(CC[n](nc1Br)nc1Br)=O</chem>	0.60	0.96
9	<chem>CC(Nc(cc1)cc(CCC2)c1C2=O)=O</chem>	1.00	0.67
10	<chem>FC(c(cc1)cc2c1[nH]c([C@@H]1CNCC1)n2)(F)F</chem>	0.79	0.79
11	<chem>CC(Nc(cc1)c(C(F)(F)F)cc1C#N)=O</chem>	0.74	0.81
12	<chem>NCc(cc1)cc2c1OCC2</chem>	0.79	0.81
13	<chem>Cc1c(CNC(c2ccccc2)=O)[s]cc1</chem>	0.76	0.85
14	<chem>OC(c1cncc(C#CCC2CCCCC2)c1)=O</chem>	1.00	0.82
15	<chem>CNCc(cc1)ccc1-c1ccc[s]1</chem>	0.80/0.86*	0.95
16	<chem>O=C(c1ccccc1)NCc(cc1)cc2c1OCC2</chem>	1.00	0.79
17	<chem>C[n](c(N)c1)nc1-c(cc1)ccc1Cl</chem>	0.76	0.73
18	<chem>O=C1NC(c(cccc2)c2F)SC1</chem>	0.76/0.78*	0.85
19	<chem>CC(C(Nc1cc(Cl)cc(Cl)c1)=O)O</chem>	0.73/0.81*	0.86
20	<chem>O=C(CCC1=O)N1c(ccc(F)c1)c1F</chem>	0.46/0.54**	0.87
21	<chem>O=C(CCC1=O)N1c1cc(Cl)cc(Cl)c1</chem>	1.00	0.84
22	<chem>CN(C)C(COc1cc2ccccc2cc1)=O</chem>	0.80/0.20**	0.79
23	<chem>O=C(c1nccccc1)c1nc2ccccc2cc1</chem>	0.83	0.88
24	<chem>N#CCC(NC1CCCCC1)=O</chem>	0.83	0.86
25	<chem>OC(c1cc(-c2cnccc2)ccc1)=O</chem>	0.73	0.84
26	<chem>NCc(cc1)cnc1-c1ccccc1</chem>	0.85	0.82
27	<chem>CC(N12)=CSC1=NC(C=O)=CC2=O</chem>	0.78	0.85
28	<chem>O=C(c(cc(cc1)F)c1N1)C1=O</chem>	0.84/0.94/0.81*	0.92
29	<chem>O=C(c(c(N1)c(cc2)Cl)c2Cl)C1=O</chem>	1.00/1.00/1.00*	0.88