Bayesian Inference Elucidates the Catalytic Competency of the SARS-CoV-2 Main Protease 3CL^pro

Evans C. Wralstad, Jessica Sayers, and Ronald T. Raines*

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

*rtraines@mit.edu

Content	Page

Additional Experimental Procedures

Materials. All chemicals and reagents were of commercial reagent grade or better and were used without further purification.

Conditions. All procedures were performed in air at ambient temperature ($\sim$22 °C) and pressure (1.0 atm) unless indicated otherwise.

Production and Purification of SARS-CoV-2 3CL^pro. Authentic 3CL^pro was produced and purified following methods described previously.[1] Briefly, BL21-Gold(DE3) competent *Escherichia coli* were transformed with the pGEX-6P-1/3CL^pro plasmid by heat shock and grown overnight at 37 °C on Luria–Bertani (Miller) agar (2% w/v) containing ampicillin (100 µg/mL). A starter culture in 1× YT medium (which contained 0.8% w/v tryptone, 0.5% w/v yeast extract, 0.25% w/v NaCl, and 100 µg/mL ampicillin) was inoculated with a single transformant colony and grown overnight at 37 °C with shaking at 250 RPM. Cultures of 1× YT medium were inoculated with the starter culture to a starting OD_{600} = 0.05 and incubated at 37 °C with shaking at 250 RPM until OD_{600} = 0.8. Gene expression was induced with isopropyl β-D-1-thiogalactopyranoside (final concentration: 0.5 mM) for 5 h at 37 °C with shaking. Cultures were pelleted by centrifugation and held at −70 °C. The induction of expression was confirmed by SDS–PAGE.

Pelleted *E. coli* cells were thawed briefly and resuspended in 20 mM Tris–HCl buffer, pH 7.8, containing NaCl (150 mM) at room temperature. *E. coli* were passed through a cell disruptor (Constant Biosystems), and the lysate was subjected to centrifugation at 38,400g and 4 °C for 2 h. The supernatant was loaded onto a HisTrap FF column (Cytiva) that had been pre-equilibrated with lysis buffer. The column was washed with 20 column-volumes of lysis buffer, and the target protein was eluted with a linear gradient (20 column volumes) of lysis buffer containing imidazole (0–500 mM). Fractions with target protein were simultaneously treated with PreScission protease (Cytiva) and dialyzed against 20 mM Tris–HCl buffer, pH 7.8, containing NaCl (150 mM) and DTT (1 mM) at 4 °C overnight with gentle stirring. The retentate was passed over coupled GSTtrap FF and HisTrap FF columns (Cytiva) to remove the GST tag, His tag, PreScission protease, and unprocessed 3CL^pro. The flowthrough was exchanged into 20 mM Tris–HCl buffer, pH 8.0, containing DTT (1 mM), loaded on a HiTrap Q FF column (Cytiva) pre-equilibrated with the same buffer, and eluted by a linear gradient of exchange buffer with 0–500 mM NaCl over 20 column volumes. Fractions with target protein were pooled and exchanged into 20 mM Tris–HCl buffer, pH 7.8, containing NaCl (150 mM), EDTA (1 mM), DTT (1 mM). Purified protein was flash-frozen in liquid nitrogen, quantified by BCA assay, and stored at −70 °C.

Analysis of 3CL^pro Purity. For Q–TOF mass spectrometry of purified 3CL^pro, a 1 µM solution of protein was made in aqueous acetonitrile (5% v/v) containing formic acid (0.1% v/v). A 15-µL aliquot (15 pmol of 3CL^pro) was injected onto a 40 °C-thermostatted PLRP-S column (5-µm particle, 1000-Å pore; Agilent Technologies) and eluted with a gradient of aqueous acetonitrile (5–95% v/v) over 7.5 min. Eluted protein was introduced to an Agilent 6530 Q–TOF mass analyzer by ESI and analyzed in positive mode.

Differential Scanning Fluorimetry. The thermostability of 3CL^pro was determined with differential scanning fluorimetry (DSF). To do so, 3CL^pro was mixed with SYPRO Orange protein gel stain (Supelco) in 50 mM HEPES–NaOH buffer, pH 7.5, to a final concentration of 100 µg/mL protein, 50× SYPRO Orange (vendor stock: 5000× in DMSO). The resulting solution was then heated from 15–95 °C at 1 °C/min using a QuantStudio 7 Flex Real-Time PCR system (Applied Biosystems); fluorescence was monitored in real-time with λ_{ex} = 470 ± 15 nm and λ_{em} = 586 ± 10 nm. Data were processed with Protein Thermal Shift software (Applied Biosystems) using the

Boltzmann fitting method. Values of T_m represent the temperature at which fluorescence was 50% maximal.

Design and Synthesis of a 3CL^pro Substrate. Aided by the homology of 3CL^pro from SARS-CoV and SARS-CoV-2, known polyprotein cleavage sites for SARS-CoV,[2] and simulated interactions in a Michaelis complex of SARS-CoV-2 3CL^pro with a peptidic substrate,[3] a candidate substrate sequence ATLQ↓SGNA (↓, cleavage site) was chosen. The octapeptide was flanked by EDANS and DABCYL conjugated to glutamic acid and lysine, respectively, to serve as a FRET pair; the peptide was further capped by terminal arginine residues to improve solubility. The desired peptide R–E(EDANS)–ATLQ↓SGNA–K(DABCYL)–R was synthesized by solid-phase peptide synthesis. Briefly, 2-chlorotrityl chloride resin loaded with Fmoc-Arg(Pbf)-OH was deprotected and coupled to Fmoc-Lys(DABCYL)-OH using PyBOP/DIPEA as the activating agent. The substrate peptide sequence was then extended using a CEM Liberty Blue automated peptide synthesizer prior to manual coupling of the Fmoc-Glu(EDANS)-OH residue, which was achieved using PyBOP/DIPEA pre-activation. Deprotection of the N-terminal Fmoc group was then performed before the final Fmoc-Arg(Pbf)-OH residue was coupled using PyBOP/DIPEA. The resulting peptide was deprotected and cleaved under standard SPPS conditions and purified by preparative reversed-phase HPLC. The control peptide R–E(EDANS)–ATLQ, which represents the N-terminal cleavage product of the substrate peptide, was synthesized by analogous means.

Analysis of Peptide Purity. For analytical HPLC of the two peptides, 50 μM peptide was dissolved in water containing trifluoroacetic acid (0.1% v/v). Peptide (2.5 nmol) was injected onto a non-thermostatted C18 column (250 × 4.6 mm, 5-μm particle, 100-Å pore; Varian) and eluted with a gradient of acetonitrile (15–40% v/v) in water over 10 min. Eluted peptide was detected with a diode array detector at λ = 210, 336, and 494 nm for peptide bond, EDANS, and DABCYL absorption, respectively.

For MALDI–TOF mass spectrometry of the two peptides, desalted peptide was spotted on an α-cyano-4-hydroxycinnamic acid matrix and analyzed with a Bruker microflex MALDI–TOF mass spectrometer in linear positive mode.

Table S1. Survey of Literature SARS-CoV and SARS-CoV-2 3CL^{pro} Steady-State Kinetic Parameters.

Virus	Enzyme[a]	[Enzyme] (nM)	Substrate[b]	pH	[NaCl] (mM)	T (°C)[c]	k_{cat} (s^{-1})	K_M (µM)	k_{cat}/K_M (M^{-1}s^{-1})	Reference
CoV-2	Native	25–500	R–E(EDANS)–ATLQ↓SGNA–K(DABCYL)–R	7.5	0	25	9.8	229	4.3×10^4	This Work
CoV-2	Native	200	MCA–AVLQ↓SGFR–K(DNP)–K	7.3	0	30	NR	NR	2.9×10^4	4
CoV-2	Native	100	(MCA)–AVLQ↓SGFR–K(DNP)–K–NH$_2$	7.3	0	27	0.54	19	2.8×10^4	5
CoV-2	Native	250	K(DABCYL)–TSAVLQ↓SGFRKM–E(EDANS)	8.0	100	25	0.52	41	1.3×10^4	6
CoV-2	+His (U)	200	K(DABCYL)–TSAVLQ↓SGFRKM–E(EDANS)-NH$_2$	8.0	150	NR	0.04	11	3.6×10^3	7
CoV-2	Native	2000	K(DABCYL)–TSAVLQ↓SGFRKM–E(EDANS)–NH$_2$	7.3	100	NR	NR	NR	3.4×10^3	1
CoV	Native	0–3000	(ARLQ↓NH)$_2$-rhodamine	8.0	0	RT/37	NR	NR	2.1×10^8	8
CoV	Native	50	K(DABCYL)–TSAVLQ↓SGFRKM–E(EDANS)	7.0	0	25	1.9	17	1.1×10^5	9
CoV	Native	25	K(DABCYL)–TSAVLQ↓SGFRKM–E(EDANS)	7.0	0	RT	1.5	45	3.4×10^4	10
CoV	Native	200	MCA–AVLQ↓SGFR–K(DNP)-K–NH$_2$	7.3	0	30	1.1	40	2.7×10^4	11
CoV	Native	200	SITSAVLQ↓SGFRKMA	7.5	0	25	8.5	600	1.4×10^4	10
CoV	Native	200	S(TAMRA)–ITSAVLQ↓SGFRKMA–K(DABCYL)	7.0	0	RT	NR	NR	1.4×10^4	10
CoV	Native	1000	K(DABCYL)–TSAVLQ↓SGFRKM–E(EDANS)	7.4	10	25	0.64	56	1.1×10^4	12
CoV	+GPH$_6$ (C)	200	MCA–AVLQ↓SGFR–K(DNP)-K–NH$_2$	7.3	0	30	0.41	61	6.8×10^3	11
CoV	+GS (N)	1000	MCA–AVLQ↓SGFR–K(DNP)-K–NH$_2$	7.3	0	30	0.14	129	1.1×10^3	13
CoV	Δ305–306	750	SWTSAVLQ↓SGFRKWA	7.0	0	25	NR	NR	1.0×10^3	14
CoV	+His (U)	1000	NMA– TSAVLQ↓SGFR–K(DNP)–M	8.0	150	37	0.0046	7	7.1×10^2	15
CoV	+His (C)	1070–17,100	TSAVLQ↓SGFRK–NH$_2$	7.3	0	RT	0.20	1150	1.8×10^2	16
CoV	+GPLGS (N)	2000	MCA–AVLQ↓SGFR–K(DNP)-K–NH$_2$	7.3	0	30	0.021	126	1.7×10^2	11
CoV	+His tag (C)	2800	TSAVLQ↓pNA	7.4	0	25	0.05	690	6.5×10^1	17

[a]Homologues are ordered by reported values of k_{cat}/K_M. His, polyhistidine tag; N and C, N- and C-terminal tags, respectively (U, tag orientation unknown). [b]Cleavage site indicated by ↓. [c]RT, room temperature (value not indicated by authors). NR, value not reported by authors.

Table S2. Survey of Literature SARS-CoV and SARS-CoV-2 3CL$^{\text{pro}}$ Dimer Dissociation Constants.

Virus	Enzyme[a]	[Enzyme] (nM)	pH	[NaCl] (mM)	T (°C)[b]	Method[c]	K_d (nM)	Reference
CoV-2	Native	25–500	7.5	0	25	Activity	16	This Work
CoV-2	Native	230–18,100	7.3	150	20	SV-AUC	2500	1
CoV	Native	28,600	7.5	120	25	SV-AUC	0.35	18
CoV	+GKFKKIVKGT (C)	28,600	7.5	120	25	SV-AUC	5.6	18
CoV	Native	5–150	7.0	0	25	Activity	15	9
CoV	+QTSITSAVLQ (N)	28,600	7.5	120	25	SV-AUC	17.2	18
CoV	+His (C)	1440–21,600	8.0	NR	20	SV-AUC	190	19
CoV	Unknown	100–3000	7.3	NR	NR	Activity	810	20
CoV	+M(−1)	0–1000	7.5	0	25	Activity	1000	2
CoV	Native	7400–29,500	7.4	100	20	SE-AUC	1300	12
CoV	Native	≤3000	8.0	0	RT/37	Activity	5200	8
CoV	Native	2700–217,000	8.0	150	NR	SAXS	5800–6800[d]	8
CoV	Native	990–25,000	7.5	75	RT	SDS–PAGE	12,700	8
CoV	+His (C)	5700; 11,400	8.0	100	RT	GFC	100,000	16
CoV	+MRGSH$_6$GSTM (U)	4000–100,000	7.5	100	25	ITC	227,000	21

[a]Homologues are ordered by reported values of K_d. His, polyhistidine tag; N and C, N- and C-terminal tags, respectively (U, tag orientation unknown). [b]RT, room temperature (value not indicated by authors). NR, value not reported by authors. [c]SV-AUC, sedimentation velocity analytical ultracentrifugation; ITC, isothermal titration calorimetry; GFC, analytical gel-filtration chromatography; SAXS, small-angle X-ray scattering; Activity, enzymatic activity; SE-AUC, sedimentation equilibrium analytical ultracentrifugation. [d]The authors analyzed their SAXS data in three ways (fitting the entire scattering curves, Guinier analysis, and fitting the normalized forward scattered intensity); each method produced an estimated K_d value.
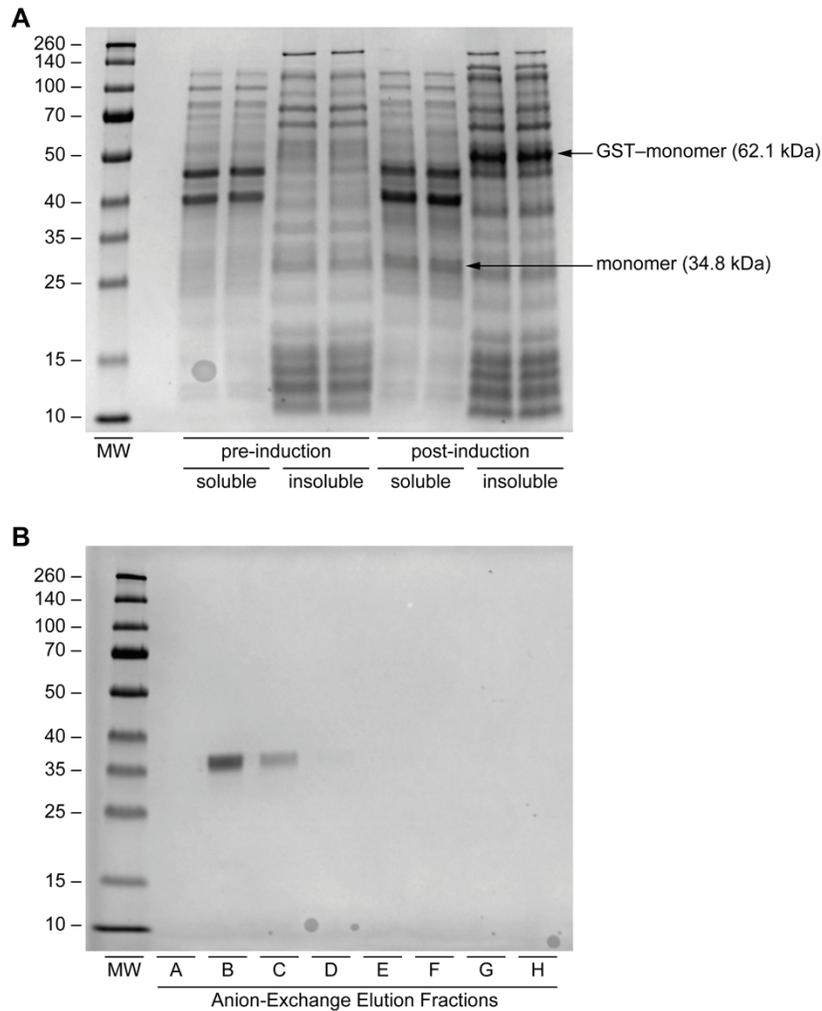
Figure S1. SDS–PAGE gels for the expression and purification of SARS-CoV-2 3CL^{pro}. (A) Expression gel. Expected masses account for a C-terminal 6× His tag. The presence of a monomer with an autolyzed N-terminal GST tag is visible in the post-induction lysate soluble fraction. The appearance of a GST-tagged monomer in the insoluble fraction of the post-induction lysate serves as a secondary confirmation of 3CL^{pro} production. (B) Purification gel. Purity of 3CL^{pro} following the removal of the His tag with PreScission protease and anion-exchange FPLC (expected mass: 33.8 kDa for 3CL^{pro} after the removal of purification tags). Elution fractions are in the order of increasing [NaCl].
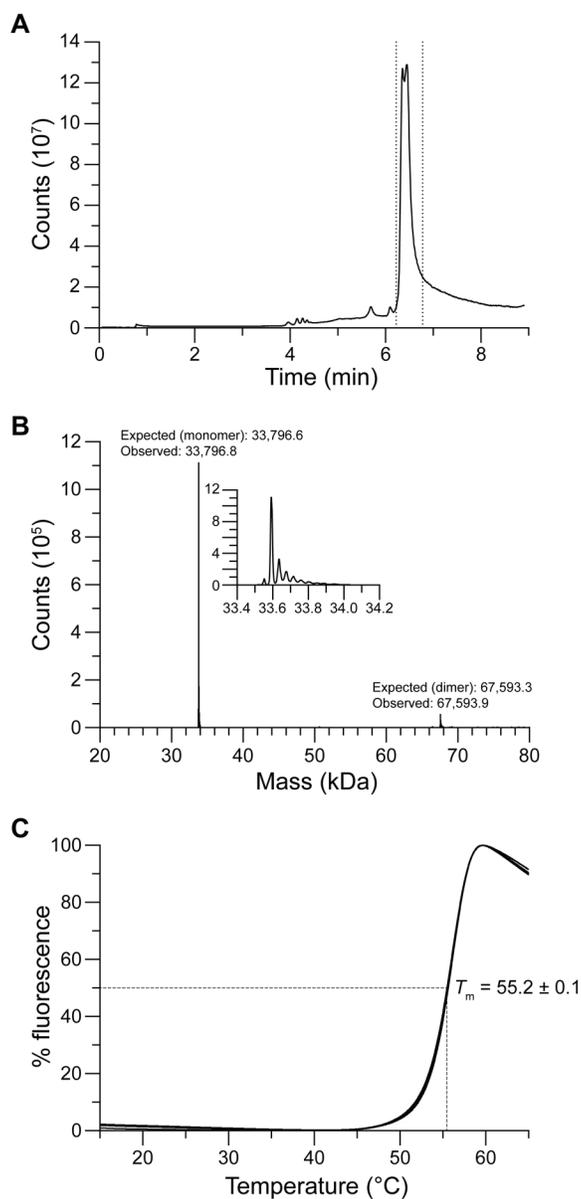
Figure S2. (A) Q–TOF total ion chromatogram of purified 3CL^{pro}. (B) Deconvoluted mass spectrum of the bounded peak in panel A. (C) DSF curves of 3CL^{pro} (100 µg/mL) in 50 mM HEPES–NaOH buffer, pH 7.5, normalized to maximum fluorescence. DSF experiments were performed in quadruplicate.
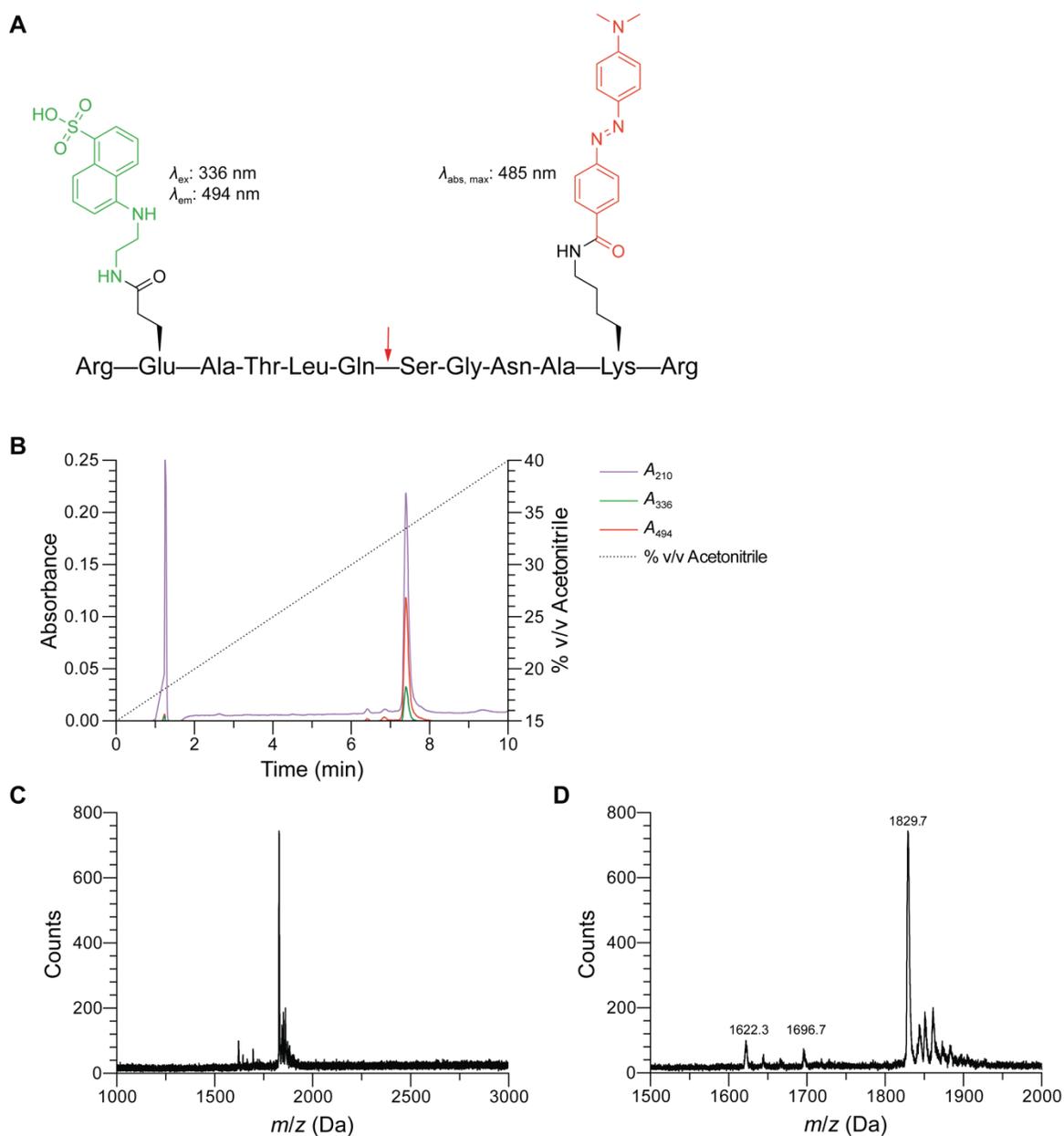
Figure S3. Characterization of the 3CL[pro] substrate R–E(EDANS)–ATLQSGNA–K(DABCYL)–R. (A) Structure of the peptide (EDANS in green; DABCYL in red). Cleavage of the Gln/Ser peptide bond by 3CL[pro] (red arrow) liberates the EDANS-bearing product peptide, increasing its fluorescence. (B) Analytical reversed-phase HPLC chromatogram of the synthesized peptide eluted from a C18 column with a 15–40% v/v acetonitrile gradient. Elution was monitored at $\lambda=$ 210, 336, and 494 nm for the absorption of peptide bonds, the EDANS moiety, and the DABCYL moiety, respectively. Following the solvent front at $t = 1$ min, a single major peak with absorption at all three wavelengths is observed at $t \approx 7.5$ min. (C) Full MALDI–TOF mass spectrum, and (D) 1500–2000 Da inset for the peptide. Expected mass [M + H]⁺, 1829.9 Da; observed mass, 1829.7 Da. Higher mass peaks correspond to salt adducts of the peptide; lower mass peaks at 1622.3 and 1696.7 Da correspond to artifactual, ionization-induced fragmentation at the EDANS and DABCYL moieties, respectively.
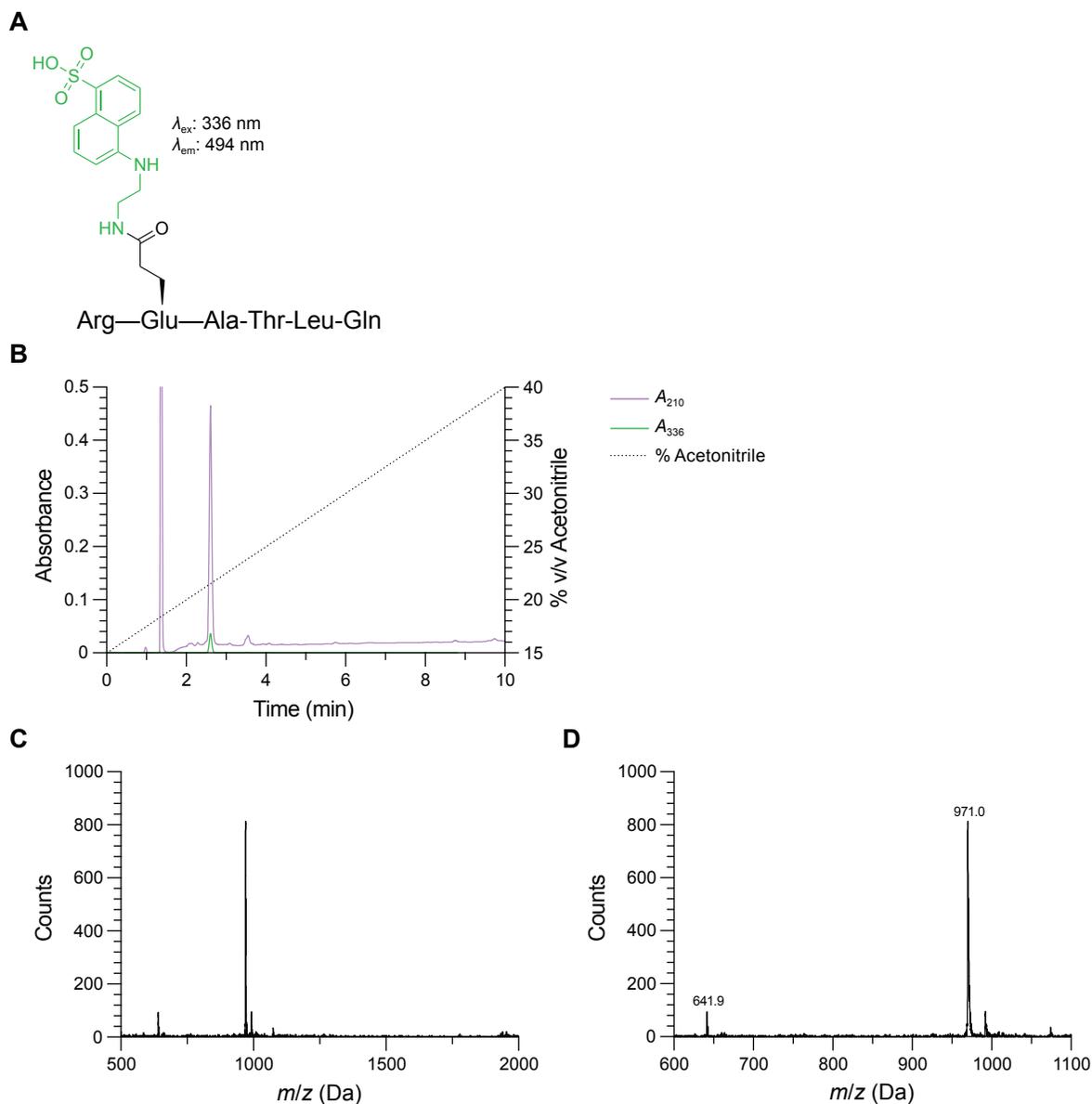
A

λ_{ex}: 336 nm
λ_{em}: 494 nm

Arg—Glu—Ala-Thr-Leu-Gln

B

C

D

Figure S4. Characterization of the 3CL^{pro} product R–E(EDANS)–ATLQ, used as a control peptide to characterize the inner filter effect (see Figure S6). (A) Structure of the peptide (EDANS in green). (B) Analytical reversed-phase HPLC chromatogram of the synthesized peptide eluted from a C18 column with a 15–40% v/v acetonitrile gradient. Elution was monitored at $\lambda = 210$ and 336 nm for the absorption of peptide bonds and the EDANS moiety, respectively. Following the solvent front at $t = 1.4$ min, a single major peak with absorption at both wavelengths is observed at $t \approx 2.6$ min. (C) Full MALDI–TOF mass spectrum, and (D) 600–1100 Da inset for the peptide. Expected mass [M + H]$^+$, 965.5 Da; observed mass, 971.0 Da. The higher mass peak corresponds to a salt adduct of the peptide; the lower mass peak at 641.9 Da corresponds to artifactual, ionization-induced fragmentation at the EDANS moiety.
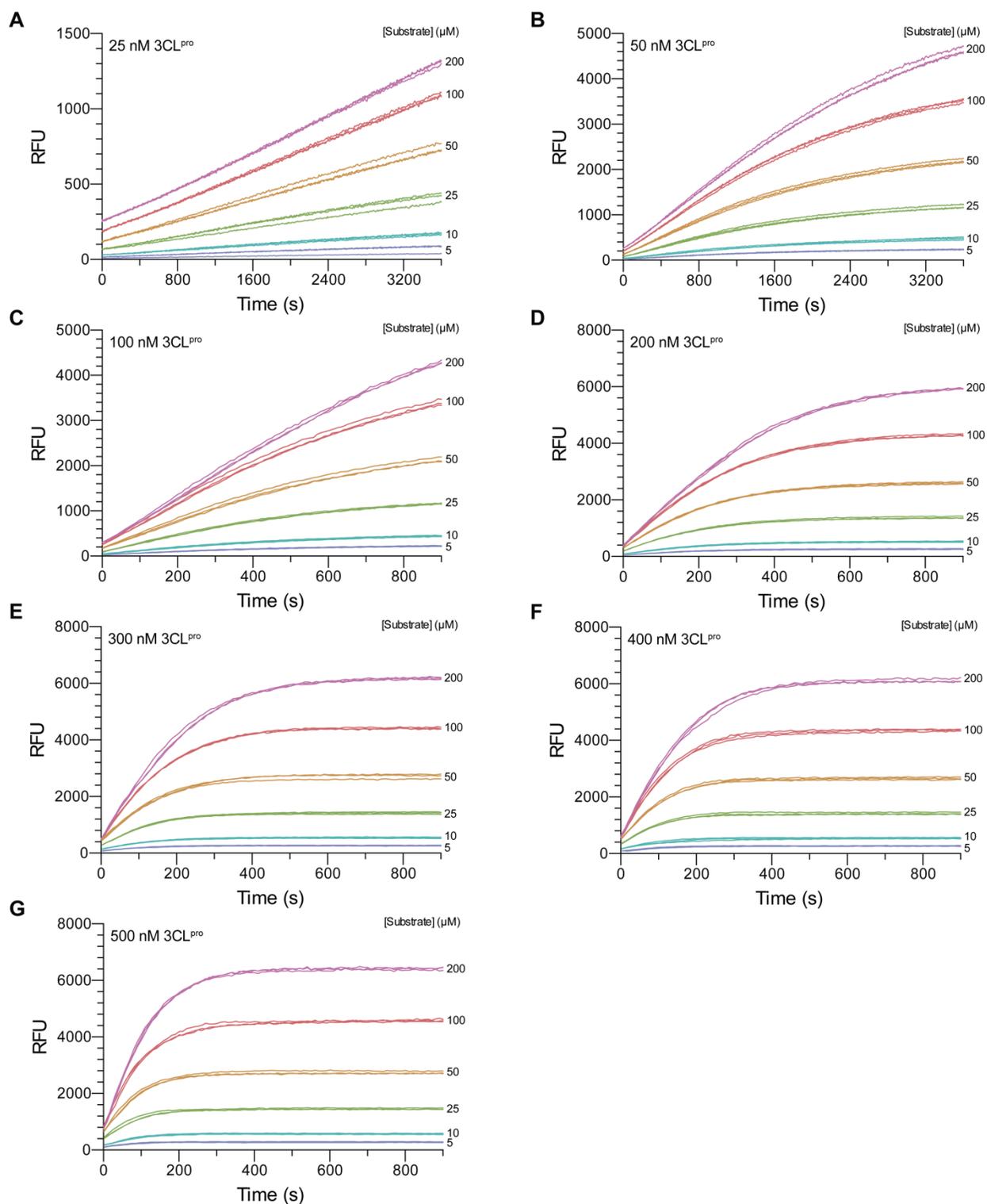
Figure S5. Progress curves for substrate cleavage by 3CL^pro at (A) 25 nM, (B) 50 nM, (C) 100 nM, (D) 200 nM, (E) 300 nM, (F) 400 nM, and (G) 500 nM. Assays in panels A and B were monitored for 60 min; assays in panels C–G were monitored for 15 min. Reactions were performed in 50 mM HEPES–NaOH buffer, pH 7.5, containing DTT (1 mM) at 25 ± 1 °C.
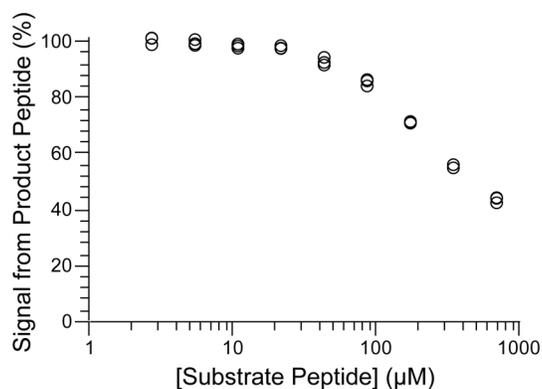
Figure S6. Graph showing the inner filter effect for the 3CL^pro substrate R–E(EDANS)–ATLQSGNA–K(DABCYL)–R. The EDANS-bearing product peptide R–E(EDANS)–ATLQ was synthesized and prepared at a final concentration of 10 µM in 50 mM HEPES–NaOH buffer, pH 7.5. The fluorescence intensity of the product peptide was measured in the presence of increasing concentrations of the substrate peptide; intensities were normalized to that of 10 µM product peptide in the absence of the substrate peptide. The attenuation of product fluorescence as a result of the inner filter effect becomes >10% as the total concentration of the substrate peptide, which contains the FRET-acceptor DABCYL, exceeds 50 µM.
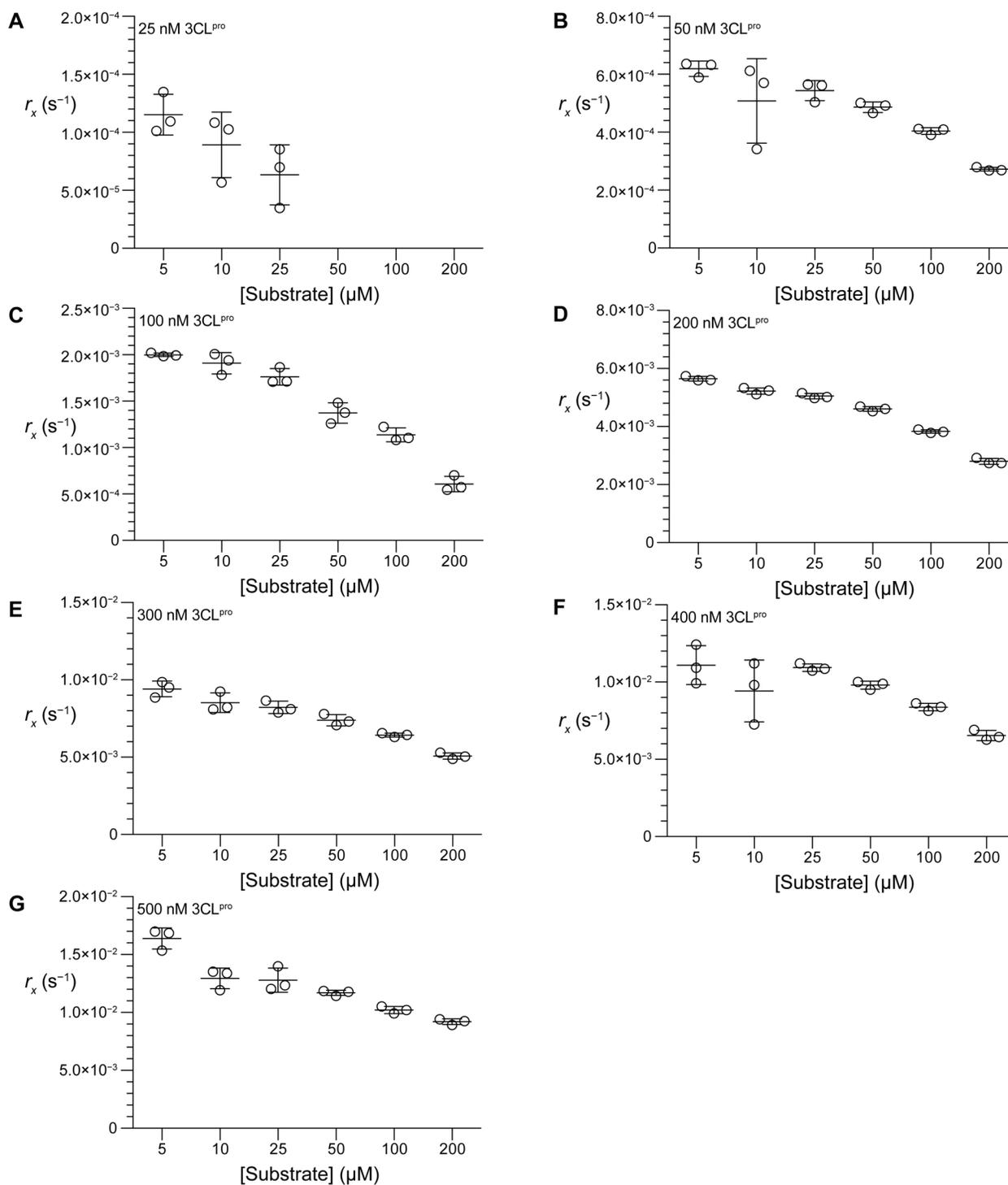
Figure S7. Fitted r_x values determined by nonlinear regression of the progress curves in Figure S5 using the pseudo-first order approximation. Values are the mean ± SD for 3CL^{pro} at (A) 25 nM, (B) 50 nM, (C) 100 nM, (D) 200 nM, (E) 300 nM, (F) 400 nM, and (G) 500 nM. Nonlinear regression failed to converge for 25 nM 3CL^{pro} at higher substrate concentrations due to the lack of a discernable plateau fluorescence intensity in the corresponding progress curves (*e.g.*, Figure S3A), resulting in a lack of fitted r_x for substrate concentrations >25 µM.
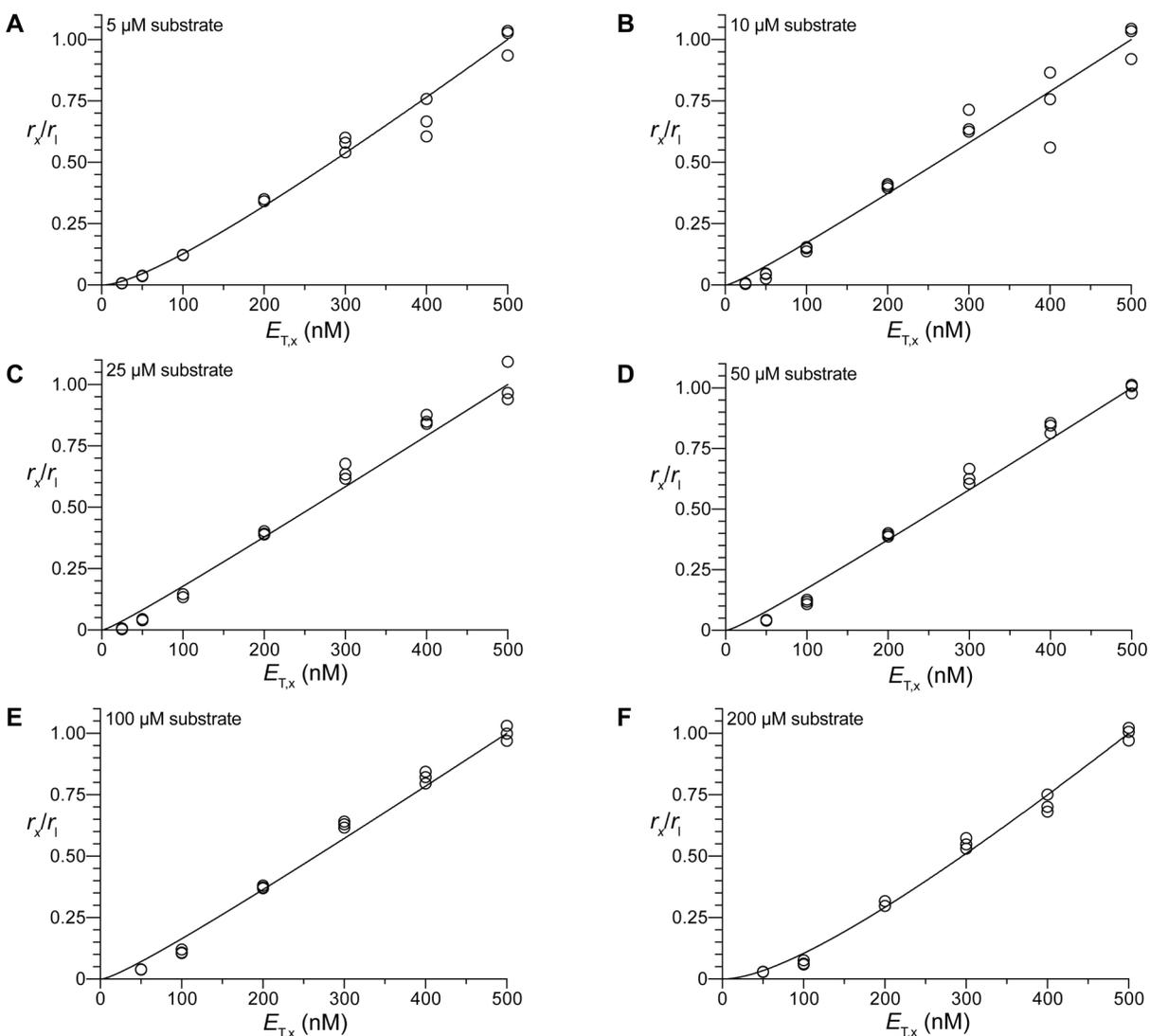
Figure S8. Fitted r_x/r_l plots to derive K_d from eq 4. Values determined by fitting the data at each substrate concentrations to eq 4 are (A) K_d = 149.2 nM at 5 µM substrate, (B) K_d = 14.87 nM at 10 µM substrate, (C) K_d = 8.782 nM at 25 µM substrate, (D) K_d = 13.59 nM at 50 µM substrate, (E) K_d = 26.52 nM at 100 µM substrate, and (F) K_d = 338.4 nM at 200 µM substrate. Only fitted K_d values for 10–100 µM substrate were used to determine the mean value of K_d because 200 µM substrate appears to violate the pseudo-first order assumption that [substrate] $\ll K_m$ (Figure S7), and the 5 µM substrate progress curves (Figure S5) have a relatively low change in fluorescence intensity over the course of the enzymatic reaction, which introduces ambiguity in nonlinear fits to eq S12.
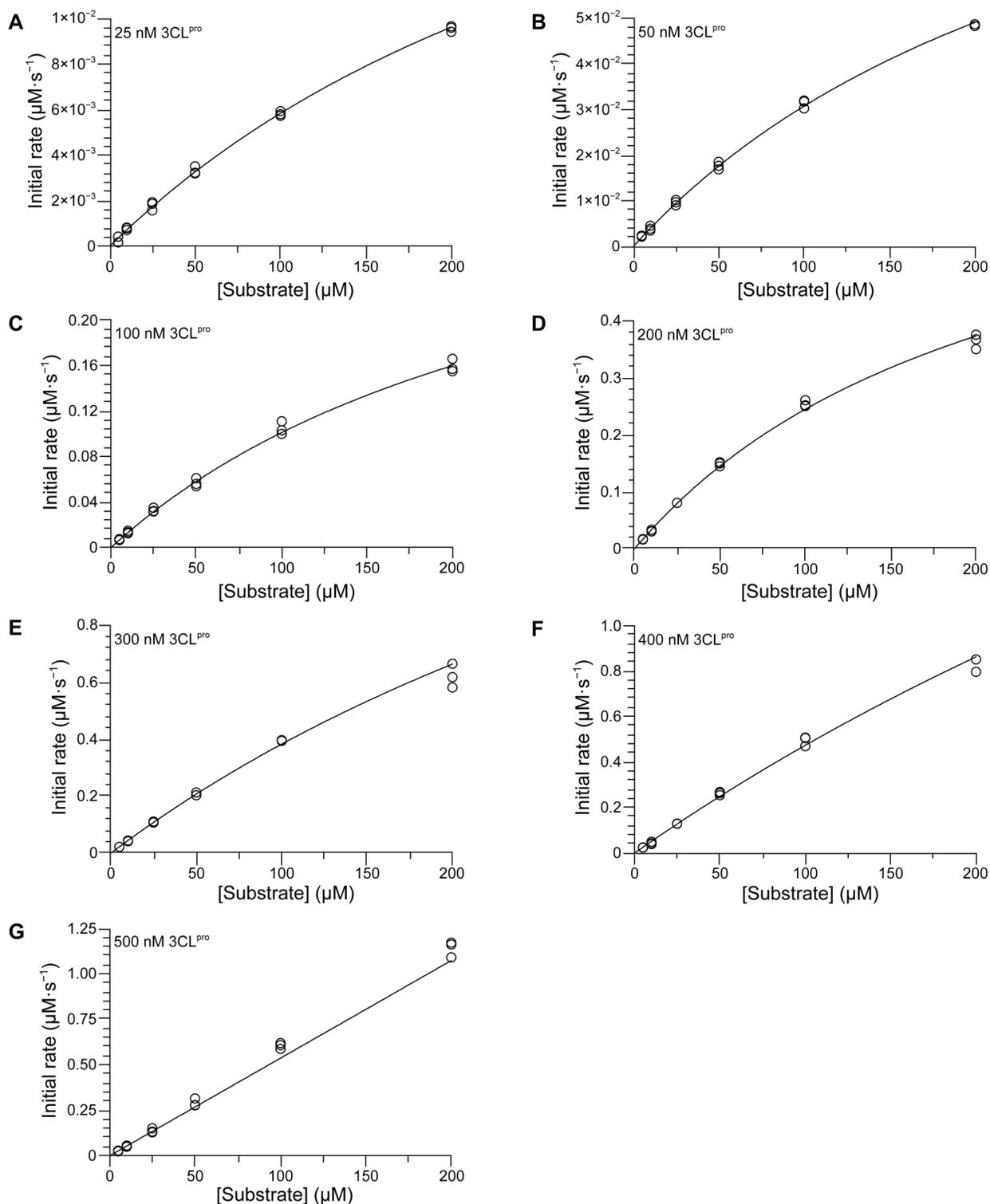
Figure S9. Michaelis–Menten curves for the cleavage of the substrate peptide by 3CL^pro at (A) 25 nM, (B) 50 nM, (C) 100 nM, (D) 200 nM, (E) 300 nM, (F) 400 nM, and (G) 500 nM. Initial rates were derived from the progress curves (Figure S5) using ICEKAT with the default setting "maximize slope magnitude."[22] Kinetic parameters are listed in Table 1.
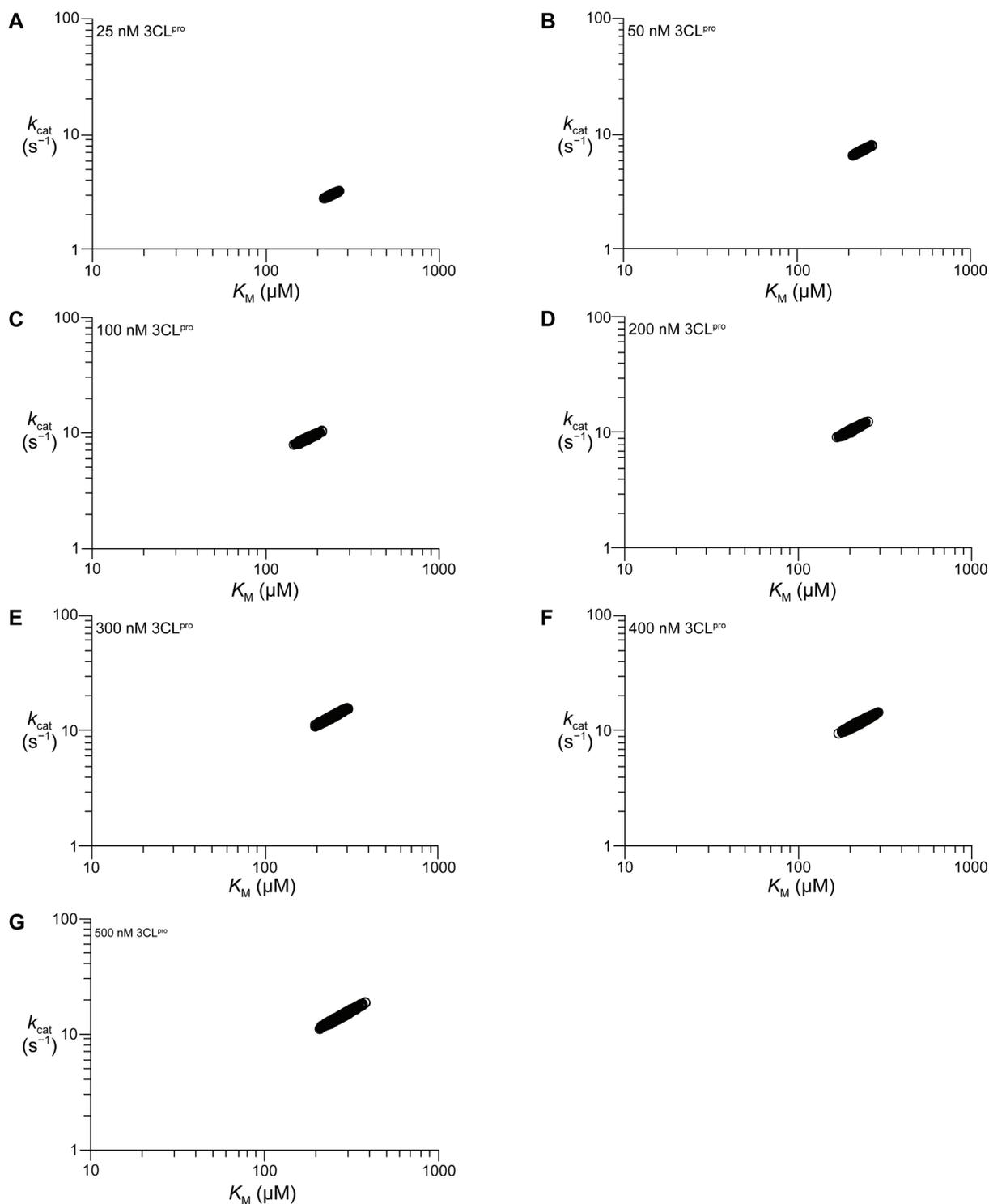
Figure S10. MCMC scatterplots of k_{cat} and K_M values ($n = 3000$ samples) for catalysis by 3CL[pro], estimated by Bayesian inference of the progress curves in Figure S5. Data were analyzed with the EKMCMC package of Hong et al. as described in the main text and using the determined 3CL[pro] dimerization K_d to convert V_{max} to k_{cat}.[23,24] MCMC scatterplots are for (A) 25 nM, (B) 50 nM, (C) 100 nM, (D) 200 nM, (E) 300 nM, (F) 400 nM, and (G) 500 nM 3CL[pro].

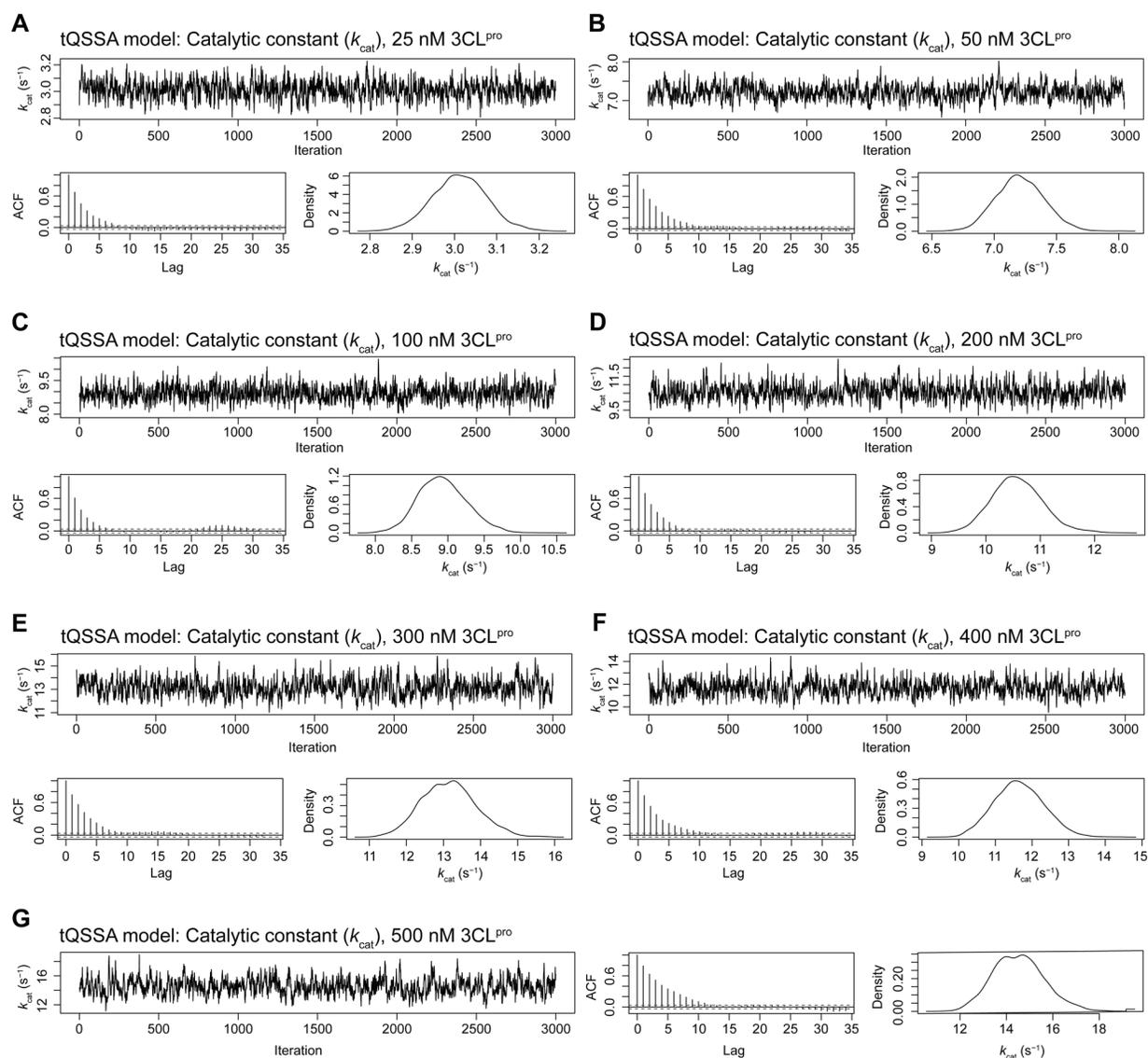**Figure S11.** Diagnostic graphs for Bayesian MCMC estimation of k_{cat} for catalysis by 3CL^{pro} from the progress curves in Figure S5, analyzed with the EKMCMC package of Hong et al. as described in the main text and using the determined 3CL^{pro} dimerization K_d to convert V_{max} to k_{cat}.[23,24] Diagnostic graphs are for (A) 25 nM, (B) 50 nM, (C) 100 nM, (D) 200 nM, (E) 300 nM, (F) 400 nM, and (G) 500 nM 3CL^{pro}. Within each sub-figure, the top graph is a trace plot showing the sampled k_{cat} values for successive iterations, the bottom-left graph is an auto-correlation function (ACF) plot showing the correlation between successive sampled k_{cat} values as a function of the distance between iterations, and the bottom-right graph is a posterior sample distribution density plot. All plots were drawn with 3000 posterior samples after removing the first 1000 samples (burn-in period) and applying a thinning rate of 1/30. All diagnostic plots are as desired: (1) the trace plots show convergent sampling, (2) the ACF plots show that successive samples rapidly become independent (ACF ≈ 0) as the lag increases, and (3) the posterior sample distribution density plots are unimodal and approximately Gaussian in shape.
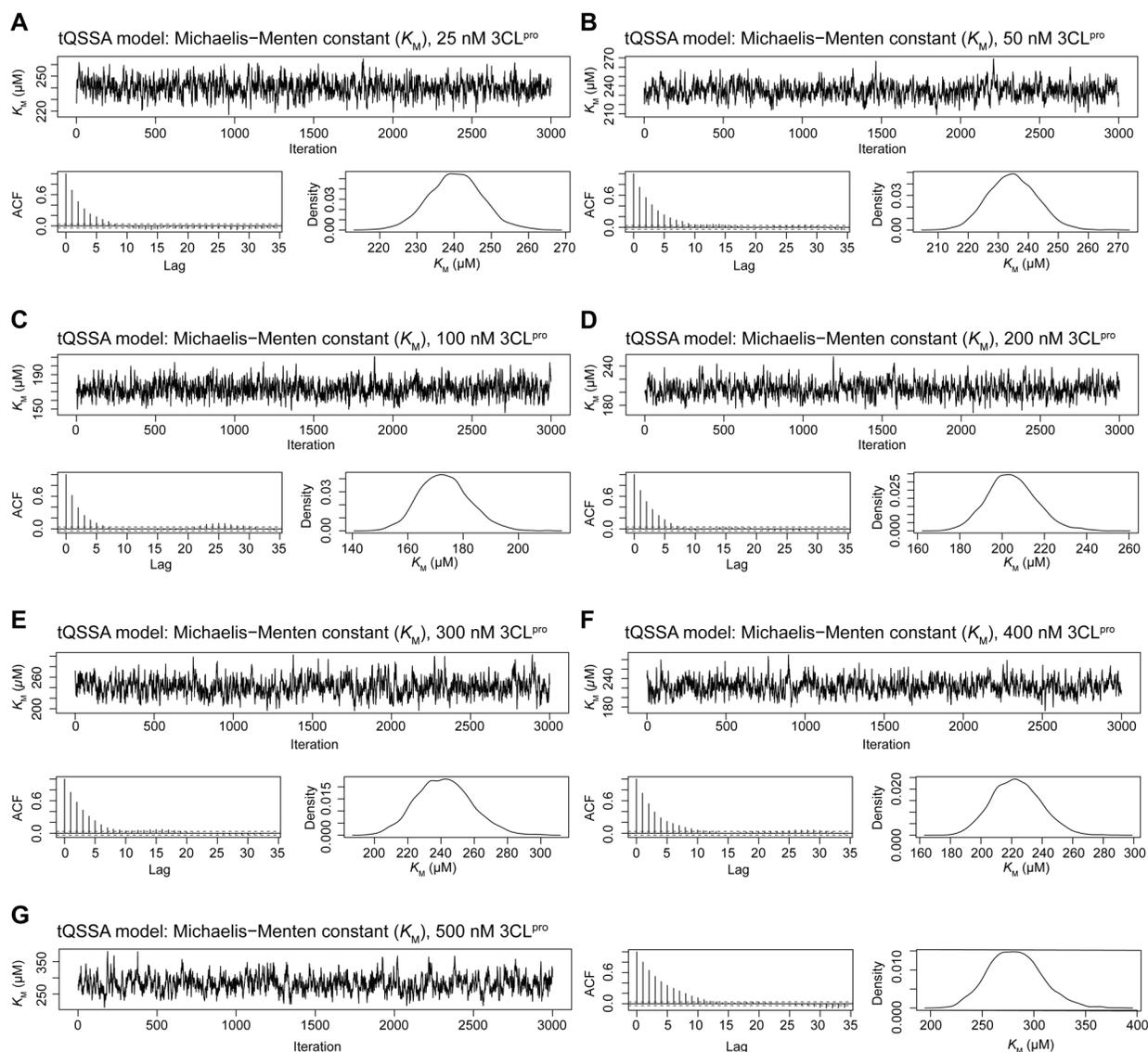
Figure S12. Diagnostic graphs for Bayesian MCMC estimation of K_M for catalysis by 3CL[pro] from progress curves in Figure S5, analyzed with the EKMCMC package of Hong et al. as described in the main text.[23-24] Diagnostic graphs are for (A) 25 nM, (B) 50 nM, (C) 100 nM, (D) 200 nM, (E) 300 nM, (F) 400 nM, and (G) 500 nM 3CL[pro]. Within each sub-figure, the top graph is a trace plot showing the sampled K_M values for successive iterations, the bottom-left graph is an auto-correlation function (ACF) plot showing the correlation between successively sampled K_M values as a function of the distance between iterations, and the bottom-right graph is a posterior sample distribution density plot. All plots were drawn with 3000 posterior samples after removing the first 1000 samples (burn-in period) and applying a thinning rate of 1/30. All diagnostic plots are as desired: (1) the trace plots show convergent sampling, (2) the ACF plots show that successive samples rapidly become independent (ACF $\approx$ 0) as the lag increases, and (3) the posterior sample distribution density plots are unimodal and approximately Gaussian in shape.

## References

(1) Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **2020**, *368*, 409–412.

(2) Grum-Tokars, V.; Ratia, K.; Begaye, A.; Baker, S. C.; Mesecar, A. D. Evaluating the 3C-like protease activity of SARS-coronavirus: Recommendations for standardized assays for drug discovery. *Virus Res.* **2008**, *133*, 63–73.

(3) Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Unraveling the SARS-CoV-2 main protease mechanism using multiscale methods. *ACS Catal.* **2020**, *10*, 12544–12554.

(4) Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of M^{pro} from COVID-19 virus and discovery of its inhibitors. *Nature* **2020**, *582*, 289–293.

(5) Lee, J.; Worrall, L. J.; Vuckovic, M.; Rosell, F. I.; Gentile, F.; Ton, A.-T.; Caveney, N. A.; Ban, F.; Cherkasov, A.; Paetzel, M.; Strynadka, N. C. J. Crystallographic structure of wild-type SARS-CoV-2 main protease acyl-enzyme intermediate with physiological C-terminal autoprocessing site. *Nat. Commun.* **2020**, *11*, 5877–5877.

(6) MacDonald, E. A.; Frey, G.; Namchuk, M. N.; Harrison, S. C.; Hinshaw, S. M.; Windsor, I. W. Recognition of divergent viral substrates by the SARS-CoV-2 main protease. *ACS Infect. Dis.* **2021**, *7*, 2591–2595.

(7) Abian, O.; Ortega-Alarcon, D.; Jimenez-Alesanco, A.; Ceballos-Laita, L.; Vega, S.; Reyburn, H. T.; Rizzuti, B.; Velazquez-Campoy, A. Structural stability of SARS-CoV-2 3CLpro and identification of quercetin as an inhibitor by experimental screening. *Int. J. Biol. Macromol.* **2020**, *164*, 1693–1703.

(8) Graziano, V.; McGrath, W. J.; Yang, L.; Mangel, W. F. SARS CoV main proteinase: The monomer–dimer equilibrium dissociation constant. *Biochemistry* **2006**, *45*, 14632–14641.

(9) Kuo, C.-J.; Chi, Y.-H.; Hsu, J. T. A.; Liang, P.-H. Characterization of SARS main protease and inhibitor assay using a fluorogenic substrate. *Biochem. Biophys. Res. Commun.* **2004**, *318*, 862–867.

(10) Solowiej, J.; Thomson, J. A.; Ryan, K.; Luo, C.; He, M.; Lou, J.; Murray, B. W. Steady-state and pre-steady-state kinetic evaluation of severe acute respiratory syndrome coronavirus (SARS-CoV) 3CL^{pro} cysteine protease: Development of an ion-pair model for catalysis. *Biochemistry* **2008**, *47*, 2617–2630.

(11) Xue, X.; Yang, H.; Shen, W.; Zhao, Q.; Li, J.; Yang, K.; Chen, C.; Jin, Y.; Bartlam, M.; Rao, Z. Production of authentic SARS-CoV M^{pro} with enhanced activity: Application as a novel Tag-cleavage endopeptidase for protein overproduction. *J. Mol. Biol.* **2007**, *366*, 965–975.

(12) Barrila, J.; Gabelli, S. B.; Bacha, U.; Amzel, L. M.; Freire, E. Mutation of Asn28 disrupts the dimerization and enzymatic activity of SARS 3CL^{pro}. *Biochemistry* **2010**, *49*, 4308–4317.

(13) Yang, H.; Xie, W.; Xue, X.; Yang, K.; Ma, J.; Liang, W.; Zhao, Q.; Zhou, Z.; Pei, D.; Ziebuhr, J.; Hilgenfeld, R.; Yuen, K. Y.; Wong, L.; Gao, G.; Chen, S.; Chen, Z.; Ma, D.; Bartlam, M.; Rao, Z. Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol.* **2005**, *3*, e324–e324.

(14) Tan, J.; Verschueren, K. H. G.; Anand, K.; Shen, J.; Yang, M.; Xu, Y.; Rao, Z.; Bigalke, J.; Heisen, B.; Mesters, J. R.; Chen, K.; Shen, X.; Jiang, H.; Hilgenfeld, R. pH-Dependent

conformational flexibility of the SARS-CoV main proteinase (M^{pro}) dimer: Molecular dynamics simulations and multiple X-ray structure analyses. *J. Mol. Biol.* **2005**, *354*, 25–40.

(15) Wang, H.; He, S.; Deng, W.; Zhang, Y.; Li, G.; Sun, J.; Zhao, W.; Guo, Y.; Yin, Z.; Li, D.; Shang, L. Comprehensive insights into the catalytic mechanism of Middle East respiratory syndrome 3C-like protease and severe acute respiratory syndrome 3C-like protease. *ACS Catal.* **2020**, *10*, 5871–5890.

(16) Fan, K.; Wei, P.; Feng, Q.; Chen, S.; Huang, C.; Ma, L.; Lai, B.; Pei, J.; Liu, Y.; Chen, J.; Lai, L. Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. *J. Biol. Chem.* **2004**, *279*, 1637–1642.

(17) Huang, C.; Wei, P.; Fan, K.; Liu, Y.; Lai, L. 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. *Biochemistry* **2004**, *43*, 4568–4574.

(18) Hsu, M.-F.; Kuo, C.-J.; Chang, K.-T.; Chang, H.-C.; Chou, C.-C.; Ko, T.-P.; Shr, H.-L.; Chang, G.-G.; Wang, A. H. J.; Liang, P.-H. Mechanism of the maturation process of SARS-CoV 3CL protease. *J. Biol. Chem.* **2005**, *280*, 31257–31266.

(19) Chou, C. Y.; Chang, H.-C.; Hsu, W.-C.; Lin, T.-Z.; Lin, C.-H.; Chang, G.-G. Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease. *Biochemistry* **2004**, *43*, 14958–14970.

(20) Chen, H.; Wei, P.; Huang, C.; Tan, L.; Liu, Y.; Lai, L. Only one protomer is active in the dimer of SARS 3C-like proteinase. *J. Biol. Chem.* **2006**, *281*, 13894–13898.

(21) Chen, S.; Chen, L.; Tan, J.; Chen, J.; Du, L.; Sun, T.; Shen, J.; Chen, K.; Jiang, H.; Shen, X. Severe acute respiratory syndrome coronavirus 3C-like proteinase N terminus is indispensable for proteolytic activity but not for enzyme dimerization. *J. Biol. Chem.* **2005**, *280*, 164–173.

(22) Olp, M. D.; Kalous, K. S.; Smith, B. C. ICEKAT: An interactive online tool for calculating initial rates from continuous enzyme kinetic traces. *BMC Bioinf.* **2020**, *21*, 186–186.

(23) Choi, B.; Rempala, G. A.; Kim, J. K. Beyond the Michaelis–Menten equation: Accurate and efficient estimation of enzyme kinetic parameters. *Sci. Rep.* **2017**, *7*, 17018–17018.

(24) Hong, H.; Choi, B.; Kim, J. K. Beyond the Michaelis–Menten: Bayesian inference for enzyme kinetic analysis. *Methods Mol. Biol.* **2022**, *2385*, 47–64.

Primer: Determining the K_d of an Obligate Homodimeric Enzyme with Half-Site Reactivity from Progress Curves

1 Michaelis-Menten kinetics in terms of dimer

For the dimerization of two enzyme monomers M to form the dimeric species D, $2M \rightleftharpoons D$ and:

$$K_d = \frac{[M]^2}{[D]} \tag{S1}$$

$$E_T = [M] + 2[D] \tag{S2}$$

where E_T is the analytical enzyme concentration on a per-protomer basis.

For a homodimeric enzyme in which only the dimer exhibits catalytic competency and the enzyme has half-site reactivity (i.e., protomer active sites alternate catalytic function), the maximal enzyme reaction rate is

$$V_{max} = k_{cat}[D] \tag{S3}$$

From Equation S3, the Michaelis-Menten equation can be written as:

$$v = \frac{V_{max}[S]}{K_M + [S]}$$
$$= \frac{k_{cat}[D][S]}{K_M + [S]} \tag{S4}$$

where $[S]$ is the free substrate concentration, k_{cat} is the enzyme's turnover rate, and K_M is the Michaelis constant. It is assumed that the total substrate concentration $[S]_0 >> E_T$, and thus $[S] \approx [S]_0$ (i.e., the amount of substrate bound in the substrate-enzyme complex is negligible at any given moment). We seek to recast Equation S4 in terms of the total enzyme concentration E_T and the enzyme dissociation constant K_d.

2 Quantifying dimer concentration

Equation S1 can be rewritten as

$$[M]^2 - K_d[D] = 0$$

Inserting Equation S2 results in:

$$(E_T - 2[D])^2 - K_d[D] = 0$$

which, upon expansion, is

$$4[D]^2 - (4E_T + K_d)[D] + E_T^2 = 0$$

Solving this quadratic expression in terms of $[D]$ leads to:

$$[D] = \frac{1}{8}\left(4E_T + K_d \pm \sqrt{K_d^2 + 8E_T K_d}\right)$$

From mass balance (Equation S2), we know that $[D] \leq \frac{1}{2}E_T$, which is the first term in the above expression for $[D]$. Thus, the additive solution for $[D]$ is not physically realizable, and therefore the concentration of dimer is equal to:

$$[D] = \frac{1}{8}\left(4E_T + K_d - \sqrt{K_d^2 + 8E_T K_d}\right) \tag{S5}$$

3 Reframing the Michaelis-Menten equation

With an analytical expression for [D] in hand, we can rewrite Equation S4:

$$v = \frac{1}{8}\left(4E_T + K_d - \sqrt{K_d^2 + 8E_T K_d}\right)\left(\frac{k_{cat}[S]}{K_M + [S]}\right) \tag{S6}$$

The second term in parentheses depends only on intrinsic enzyme properties (k_{cat}, K_M) and the design of the assay ([S]), and it is entirely independent of E_T.

For the first term in parentheses, we consider two asymptotic conditions. In the first case of negligble enzyme, $E_T << K_d \rightarrow 4E_T + K_d \approx K_d$ and $K_d^2 + 8K_d E_T \approx K_d^2$. As a result, v tends to zero, as is expected under conditions where nearly all of the enzyme is in the inactive monomeric state.

In the second case of abundant enzyme, $E_T >> K_d$, the parenthetical expression tends to $\frac{1}{2}E_T$, as expected under conditions where all of the enzyme is driven to dimerize.

4 Pseudo-first order conditions

Focusing on the Michaelis-Menten equation in the form described by Equation S4, with [D] described by Equation S5, we consider the case where $[S] << K_M$. Under these conditions, the Michaelis-Menten equation simplifies to:

$$v = -\frac{d[S]}{dt} = \left(\frac{k_{cat}}{K_M}\right)[D][S]$$

Integrating this equation to solve for [S] results in:

$$[S] = Ce^{-\left(\frac{k_{cat}}{K_M}\right)[D]t} \tag{S7}$$

where C is a constant of integration that is solved for upon considering the precise format of the enzyme assay (see below).

5 Fluorescence-based assay monitoring

For the catalyzed reaction $S \rightarrow P$, product concentration [P] is related to substrate concentration by $[P] = [S]_0 - [S]$ (again, it is assumed that $E_T << [S]_0$, and therefore the concentration of enzyme-bound substrate may be neglected). Substituting Equation S7 into this expression results in:

$$[P] = [S]_0 - Ce^{-\left(\frac{k_{cat}}{K_M}\right)[D]t}$$

As expected, $\lim_{t\rightarrow\infty}[P] = [S]_0$. Invoking the initial condition that $[P] = 0$ at $t = 0$ results in $C = [S]_0$, and therefore

$$[P] = [S]_0\left(1 - e^{-\left(\frac{k_{cat}}{K_M}\right)[D]t}\right) \tag{S8}$$

For an enzyme assay in which a non-fluorescent substrate is converted to a fluorescent product (such as a FRET cleavage assay), the fluorescence intensity F is related to [P] by

$$F = F_0 + \alpha[P] \tag{S9}$$

where F_0 is the background fluorescence intensity and α is a proportionality constant. Given that the maximum fluorescence intensity F_{max} will be observed when all substrate has been converted to product (i.e., $[P] = [S]_0$),

$$F_{max} = F_0 + \alpha[S]_0 \tag{S10}$$

Rearranging Equations S9 and S10 to isolate [P] and $[S]_0$, respectively, then inserting the expressions into Equation S8 results in:

$$F = F_0 + (F_{max} - F_0)\left(1 - e^{-\left(\frac{k_{cat}}{K_M}\right)[D]t}\right) \tag{S11}$$

6 Determining K_d and $\frac{k_{cat}}{K_M}$

To start, we assay the homodimeric enzyme under a range of substrate and enzyme concentrations. The substrate concentration should span orders of magnitude with the goal of assaying concentrations less than the (unknown) K_M, and the enzyme concentration E_T should be significantly less than $[S]_0$. Then, the fluorescence vs time data are fitted by nonlinear regression to a modified form of Equation S11:

$$F = F_0 + (F_{max} - F_0)\left(1 - e^{-r_x t}\right) \tag{S12}$$

where

$$r_x = \left(\frac{k_{cat}}{K_M}\right)[D]_x \tag{S13}$$

and $[D]_x$ is the concentration of enzyme dimer for a given analytical enzyme concentration $E_T = E_{T,x}$ (x denoting that several concentrations are assayed in the complete dataset). Note that r_x is independent of $[S]_0$, and for a single $E_{T,x}$ assayed across a range of $[S]_0$, a consistent fitted r_x should be determined as long as the condition $[S]_0 << K_M$ is valid. In fact, by examining at what $[S]_0$ the fitted r_x appears to begin deviating, we can qualitatively judge at what $[S]_0$ the pseudo-first order approximation begins breaking down and therefore in what approximate regime the value of the unknown K_M must be.

Focusing on a single $[S]_0$ for which the pseudo-first order approximation is valid, we determine the fitted r_x for all assayed $E_{T,x}$. Then, we let r_1 represent the fitted r_x for the largest enzyme concentration assayed, $E_{T,1}$. We normalize the fitted r_x values to r_1:

$$\frac{r_x}{r_1} = \frac{\left(\frac{k_{cat}}{K_M}\right)[D]_x}{\left(\frac{k_{cat}}{K_M}\right)[D]_1}$$

$$= \frac{[D]_x}{[D]_1}$$

Inserting Equation S5 into the above equation yields:

$$\frac{r_x}{r_1} = \frac{4E_{T,x} + K_d - \sqrt{K_d^2 + 8K_d E_{T,x}}}{4E_{T,1} + K_d - \sqrt{K_d^2 + 8K_d E_{T,1}}} \tag{S14}$$

A plot of r_x/r_1 vs $E_{T,x}$ is a function with one fittable parameter, K_d (recall that $E_{T,1}$ is simply the largest enzyme concentration used and is therefore an empirically-defined constant). Thus, nonlinear regression of the data to Equation S14 will determine K_d.

7 Summary

We assay an obligate homodimeric enzyme with half-site reactivity using a FRET pair-tagged substrate to produce a fluorescent product under a range of substrate concentrations $[S]_0$ and analytical enzyme concentrations $E_{T,x}$. We fit the data of each assay (i.e., one $[S]_0$ and $E_{T,x}$) to Equation S12:

$$F = F_0 + (F_{max} - F_0)\left(1 - e^{-r_x t}\right) \tag{S15}$$

This yields a collection of fitted r_x.

For a given $E_{T,x}$, we assess the fitted r_x across the assayed substrate concentrations and qualitatively decide where the assumption $[S]_0 << K_M$ breaks down. We restrict further analysis to the $[S]_0$ below this cutoff. Within each $[S]_0$ for which the pseudo-first order approximation appears valid, we normalize the fitted r_x to r_1, which is simply the fitted r_x at the largest enzyme concentration $E_{T,1}$. Plotting $\frac{r_x}{r_1}$ against $E_{T,x}$ and fitting the transformed data to Equation S14

$$\frac{r_x}{r_1} = \frac{4E_{T,x} + K_d - \sqrt{K_d^2 + 8K_d E_{T,x}}}{4E_{T,1} + K_d - \sqrt{K_d^2 + 8K_d E_{T,1}}} \tag{S16}$$

will yield K_d.