Sensitive detection of SARS-CoV-2 main protease 3CL^{pro} with an engineered ribonuclease zymogen

Evans C. Wralstad | Ronald T. Raines

[1]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

TABLE OF CONTENTS

TABLE S1 Location of potential off-target 3CL^{pro} cleavage sites in the zymogen and identified 3CL^{pro} substrate incompatibilities

Gln[a]	Incompatibility with 3CL^{pro} Substrate Specificity[b]
9	P4 = Lys6 (*large*), P1′ = Arg10 (*bulky, cationic*)
11	P4 = Phe8 (*bulky*), P2 = Arg10 (*cationic*), P1′ = His12 (*bulky*)
28	P4 = Tyr25 (*bulky*), P1′ = Met29 (*bulky*)
37	P2 = Thr36 (*hydrophilic*)
55	P1′ = Asn56 (*bulky*)
60	P1′ = Glu61 (*anionic*)
69	P4 = Lys66 (*bulky*)

[a]Using native RNase 1 residue numbering, the zymogen sequence is

```
5   10          20          30          40          50          60          70
AKKFQR QHMDSDSSPS SSSTYCNQMM RRRNMTQGRC KPVNTFVHEP LVDVQNVCFQ EKVTCKNGQG
         80          90         100         110         120 122 ********
NCYKSNSSMH ITDCRLTNGS RYPNCAYRTS PKERHIIVAC EGSPYVPVHF  DA ATLQSGNA
```

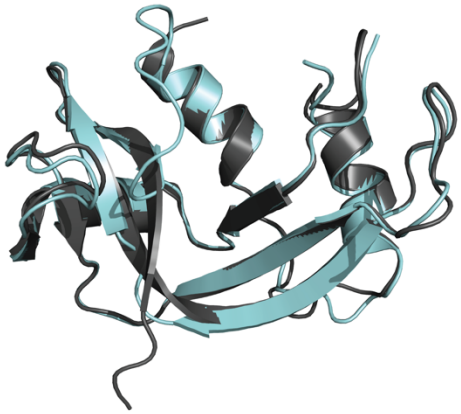
The linker is in red font and denoted by *.

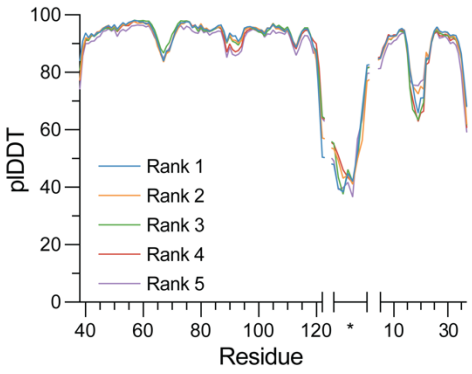
[b]Using Schechter–Berger nomenclature, 3CL^{pro} strictly requires glutamine at P1, and preferentially binds hydrophobic residues at P2; small residues at P4; and small, uncharged residues at P1′ (Anand et al., 2003; Kneller et al., 2020; Ramos-Guzmán et al., 2020; Koudelka et al., 2021).

A

```
  40          50          60          70          80          90          100         110         120
GRC KPVNTFVHEP LVDVQNVCFQ EKVTCKNGQG NCYKSNSSMH ITDCRLTNGS RYPNCAYRTS PKERHIIVAC EGSPYVPVHF

122 ******** 5   10          20          30      37
 DA ATLQSGNA AKKFQR QHMDSDSSPS SSSTYCNQMM RRRNMTQ
```

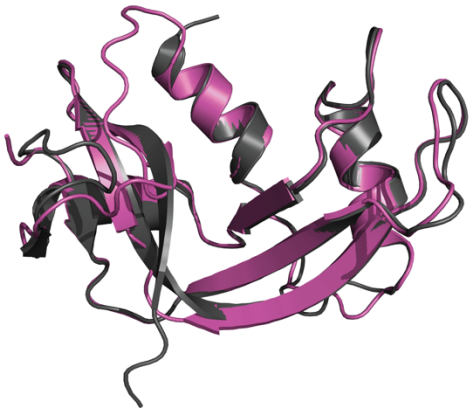


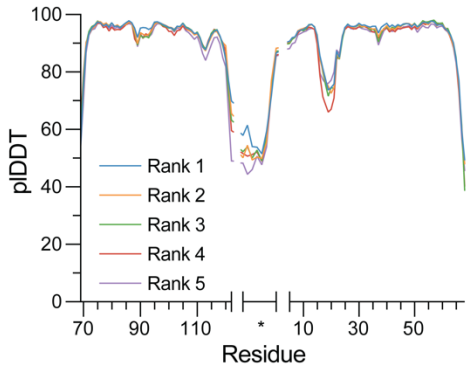


B

```
70          80          90          100         110         120 122 ******** 5   10
QG NCYKSNSSMH ITDCRLTNGS RYPNCAYRTS PKERHIIVAC EGSPYVPVHF  DA ATLQSGNA AKKFQR

        20          30          40          50          60      68
QHMDSDSSPS SSSTYCNQMM RRRNMTQGRC KPVNTFVHEP LVDVQNVCFQ EKVTCKNG
```





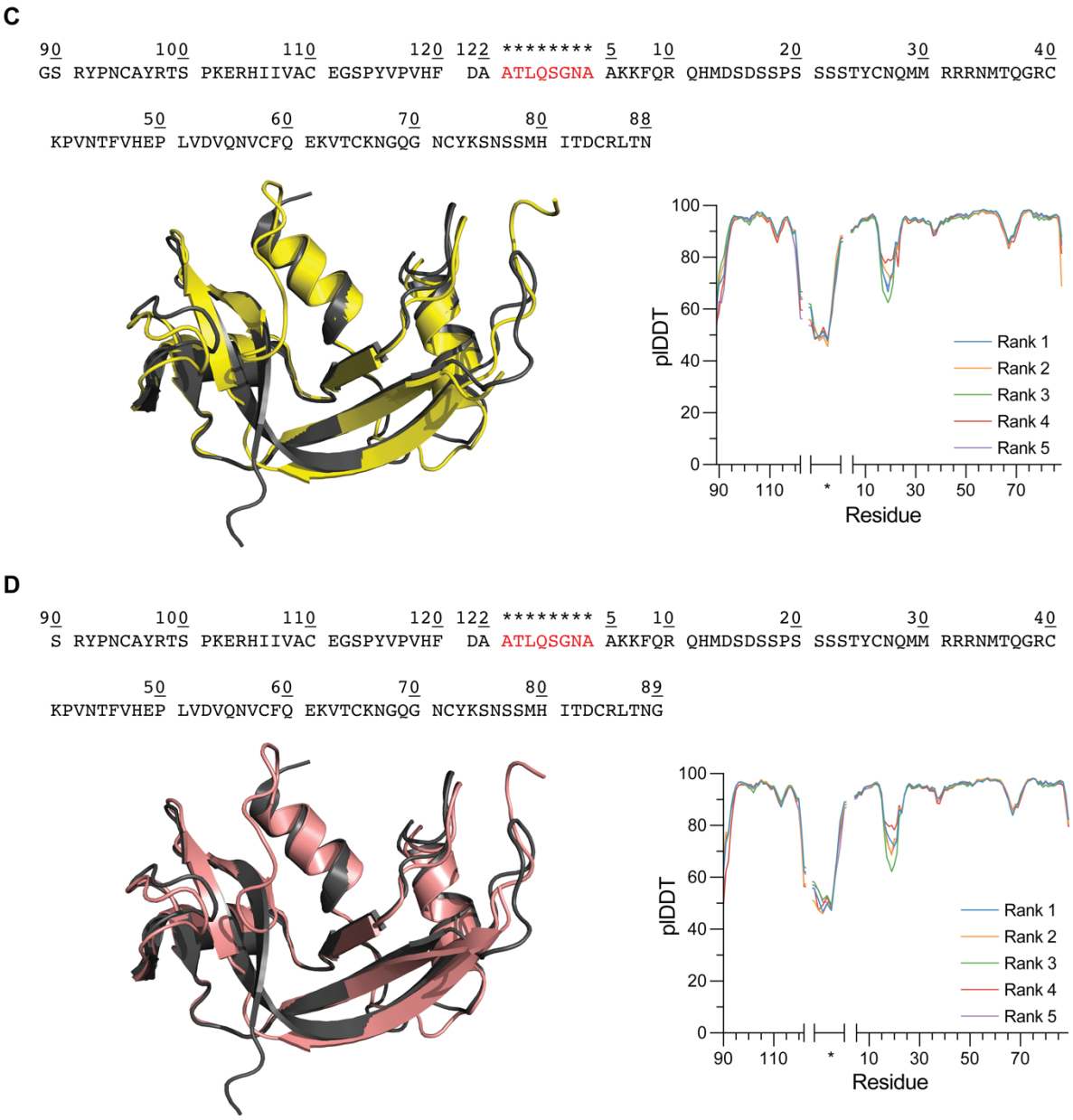


FIGURE S1 Zymogen structure predictions generated using AlphaFold2 with protein sequences circularly permuted around (A) Q37|R38, (B) G68|Q69, (C) N88|R89, and (D) R89|R90. Each subpanel shows (top) the permuted protein sequence with the linker in red font and denoted by *; (left) the top-ranked prediction model structurally aligned to wild-type RNase 1 (in gray); and (right) plDDT scores for the five models produced from a given permutant sequence.

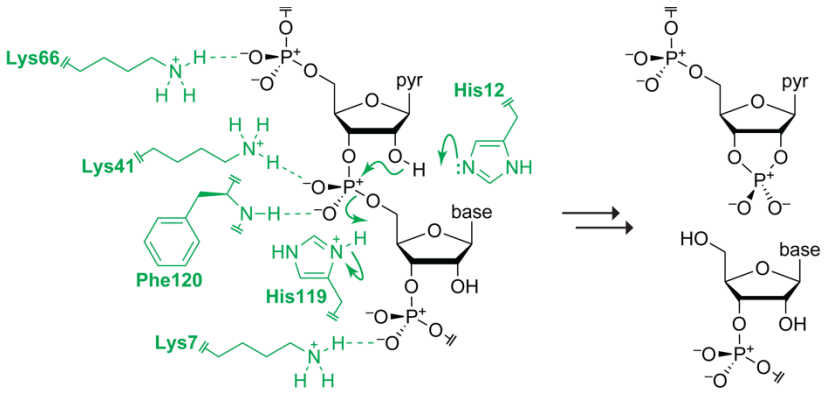
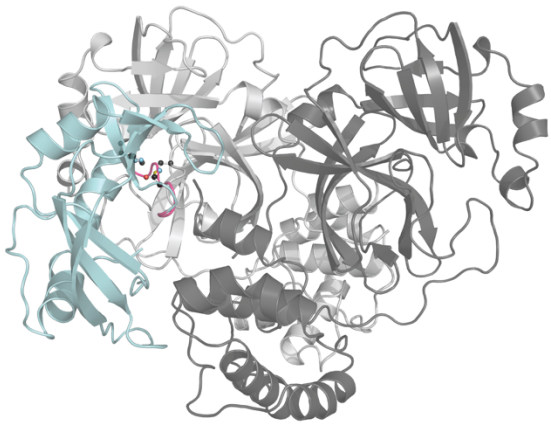
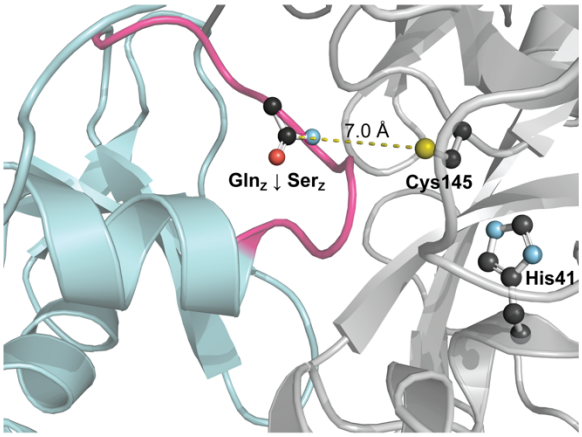


FIGURE S2 Putative mechanism of catalysis of RNA cleavage by human ribonuclease 1.

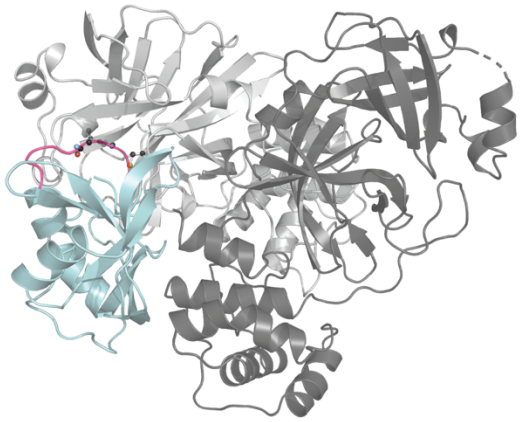
A

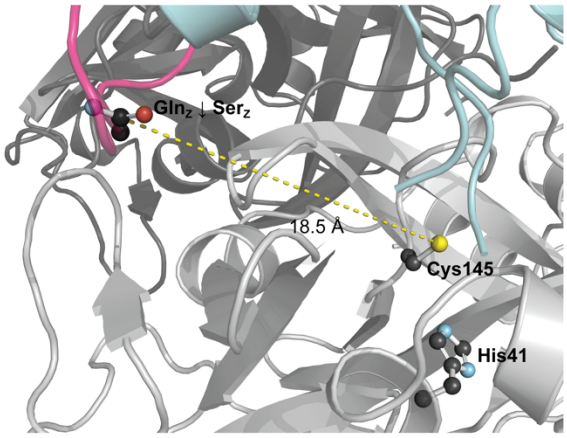




Model 3 of 30
59 members; center score: −706.4; lowest score: −808.7

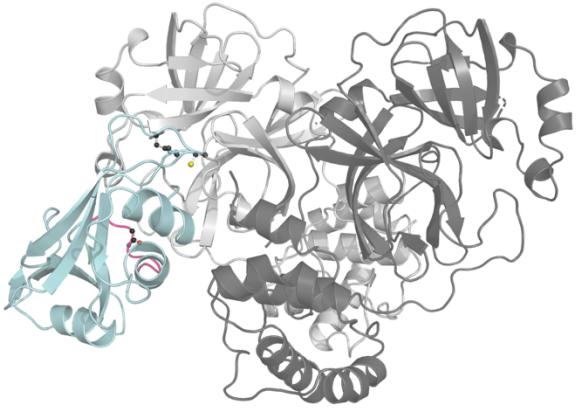
B

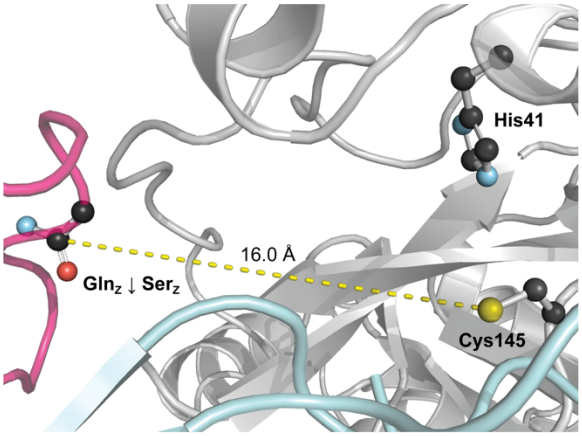




Model 10 of 30
27 members; center score: −664.2; lowest score: −821.0

C





Model 12 of 30
25 members; center score: −699.0; lowest score: −755.7

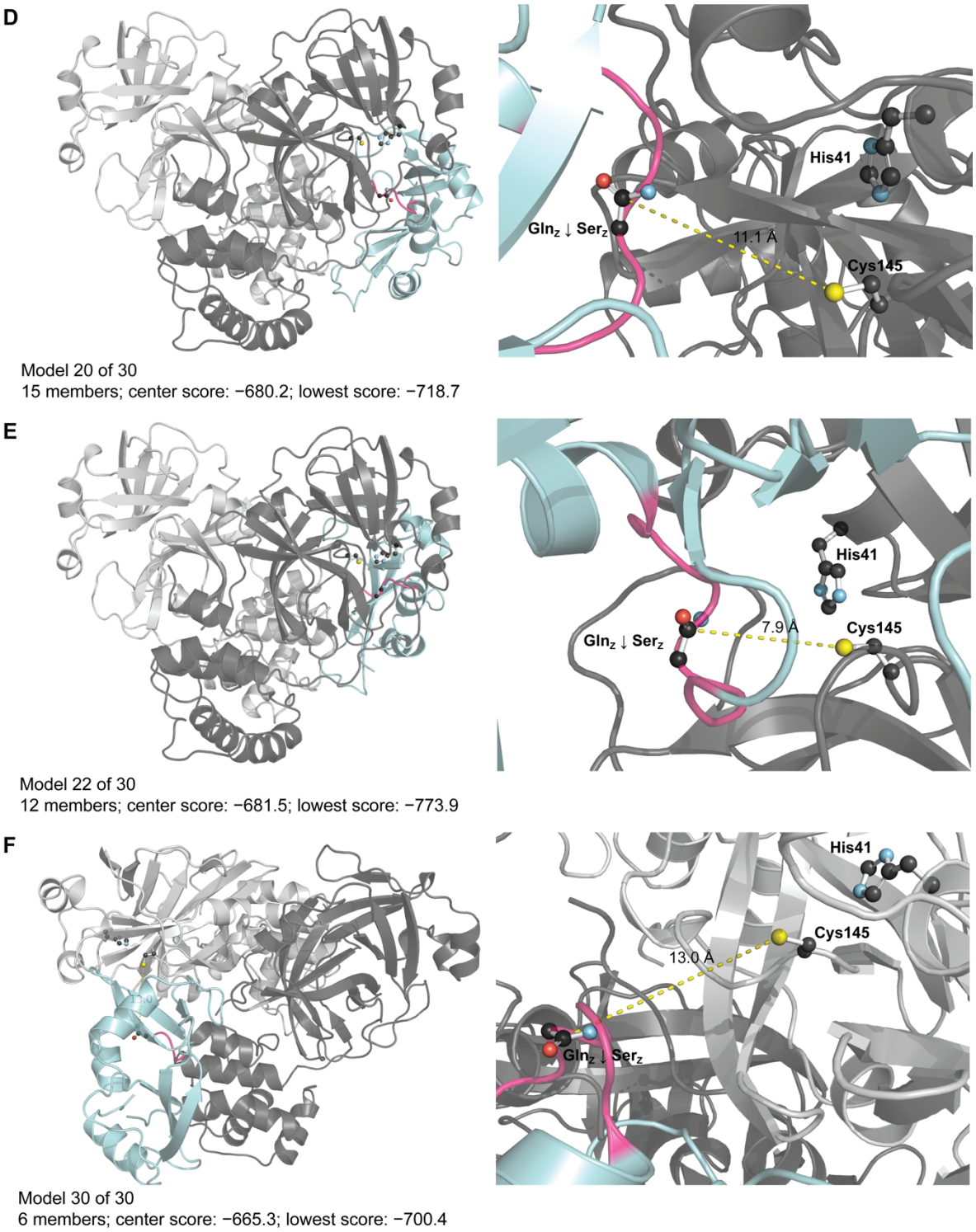
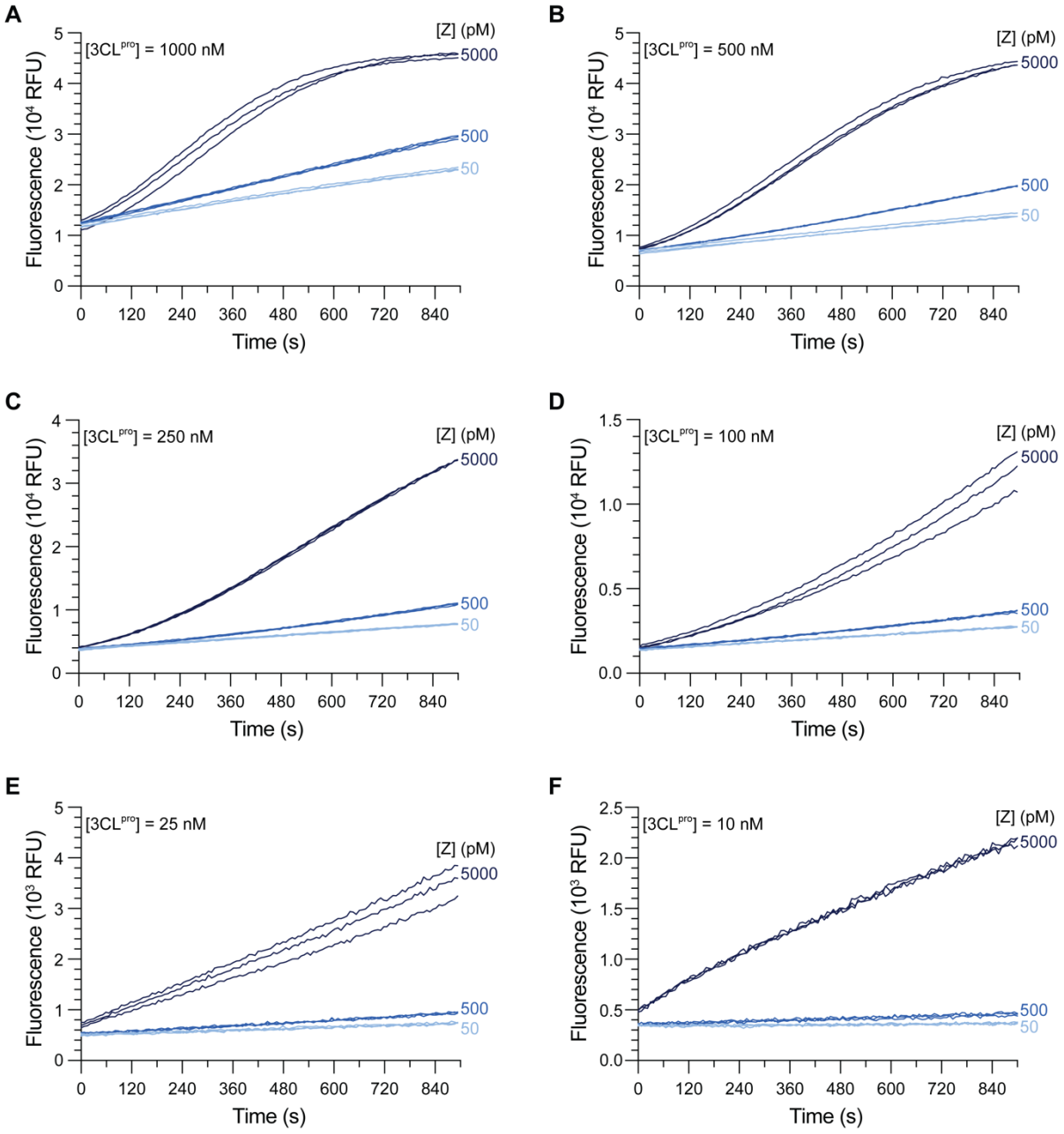


FIGURE S3 Images of ClusPro-predicted zymogen·3CL^{pro} models (A–F) in which the zymogen linker is proximal to the protease active site. Each subpanel shows (left) the global interaction of the 3CL^{pro} protomers (cartoon, distinct shades of gray) and the zymogen (cartoon, teal with linker in magenta) with atoms relevant to linker cleavage in ball-and-stick form (CPK color scheme); (right) the location of the scissile amide bond of the zymogen linker (Gln$_Z$↓Ser$_Z$) in relation to the 3CL^{pro} catalytic residue Cys145.


A
[3CL^{pro}] = 1000 nM
[Z] (pM)
5000
500
50
Fluorescence (10^4 RFU)
Time (s)
B
[3CL^{pro}] = 500 nM
[Z] (pM)
5000
500
50
Fluorescence (10^4 RFU)
Time (s)
C
[3CL^{pro}] = 250 nM
[Z] (pM)
5000
500
50
Fluorescence (10^4 RFU)
Time (s)
D
[3CL^{pro}] = 100 nM
[Z] (pM)
5000
500
50
Fluorescence (10^4 RFU)
Time (s)
E
[3CL^{pro}] = 25 nM
[Z] (pM)
5000
500
50
Fluorescence (10^3 RFU)
Time (s)
F
[3CL^{pro}] = 10 nM
[Z] (pM)
5000
500
50
Fluorescence (10^3 RFU)
Time (s)

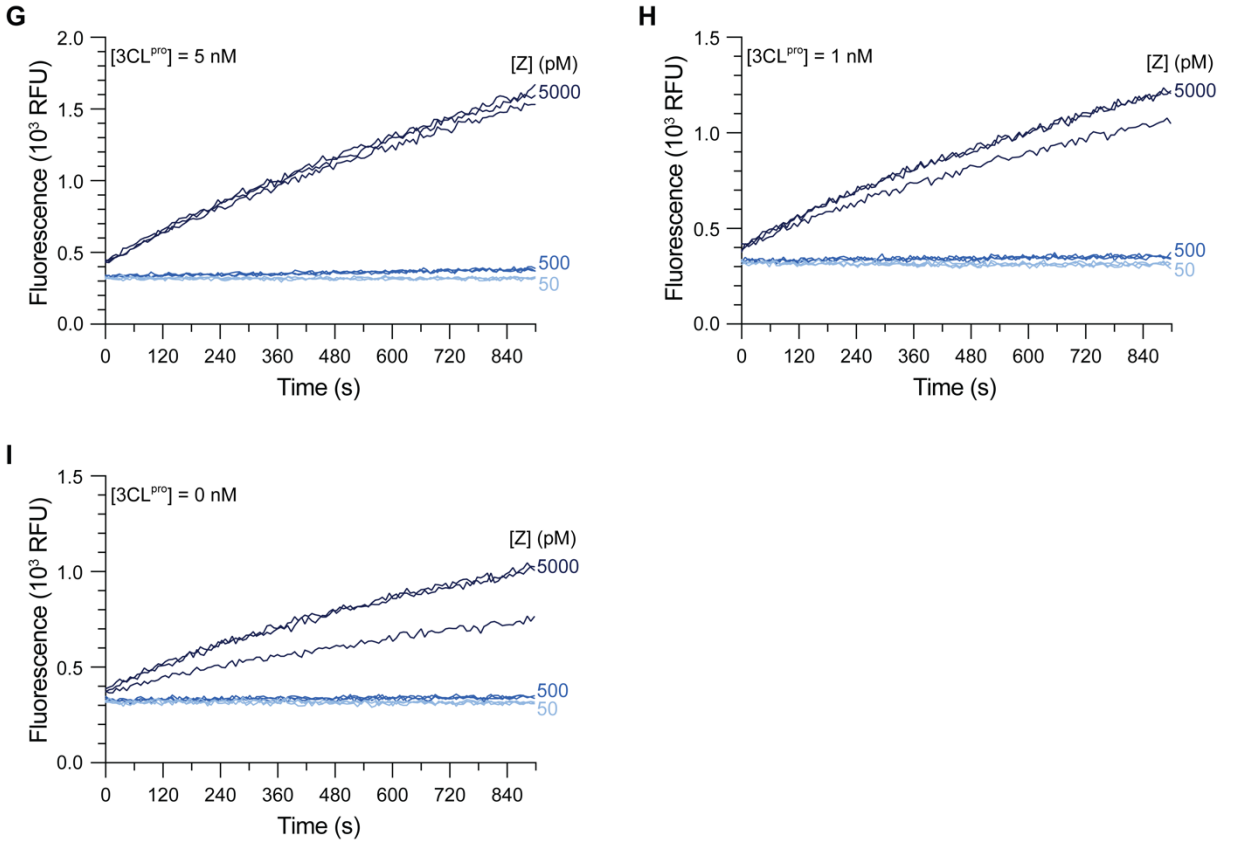


FIGURE S4 Graphs of the real-time detection of 3CL^{pro} via zymogen activation and subsequent ribonucleolysis. Each panel depicts zymogen activity at one [3CL^{pro}]: (A) 1000 nM; (B) 500 nM, (C) 250 nM, (D) 100 nM, (E) 25 nM, (F) 10 nM, (G) 5 nM, (H) 1 nM, and (I) 0 nM. All reactions were conducted with 200 nM substrate in 100 mM Tris–HCl buffer, pH 7.5, containing NaCl (10 mM) at 25 °C.

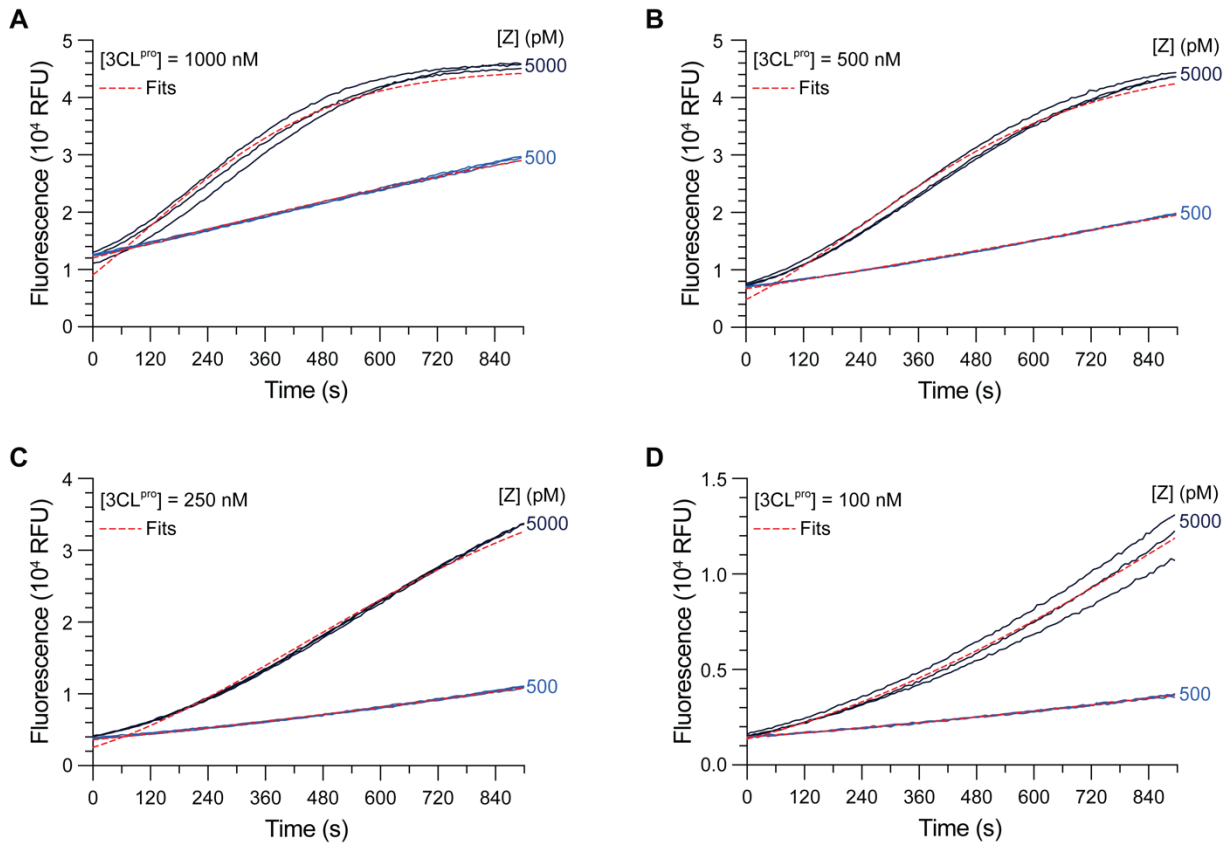


FIGURE S5 Graphs of the fitting of data from Figure S4 for one [3CL^{pro}]. (A) 1000 nM, (B) 500 nM, (C) 250 nM, and (D) 100 nM. Catalytic efficiencies were determined by nonlinear regression of each time-course data set to eq S14; each fit is indicated by a red dashed line.

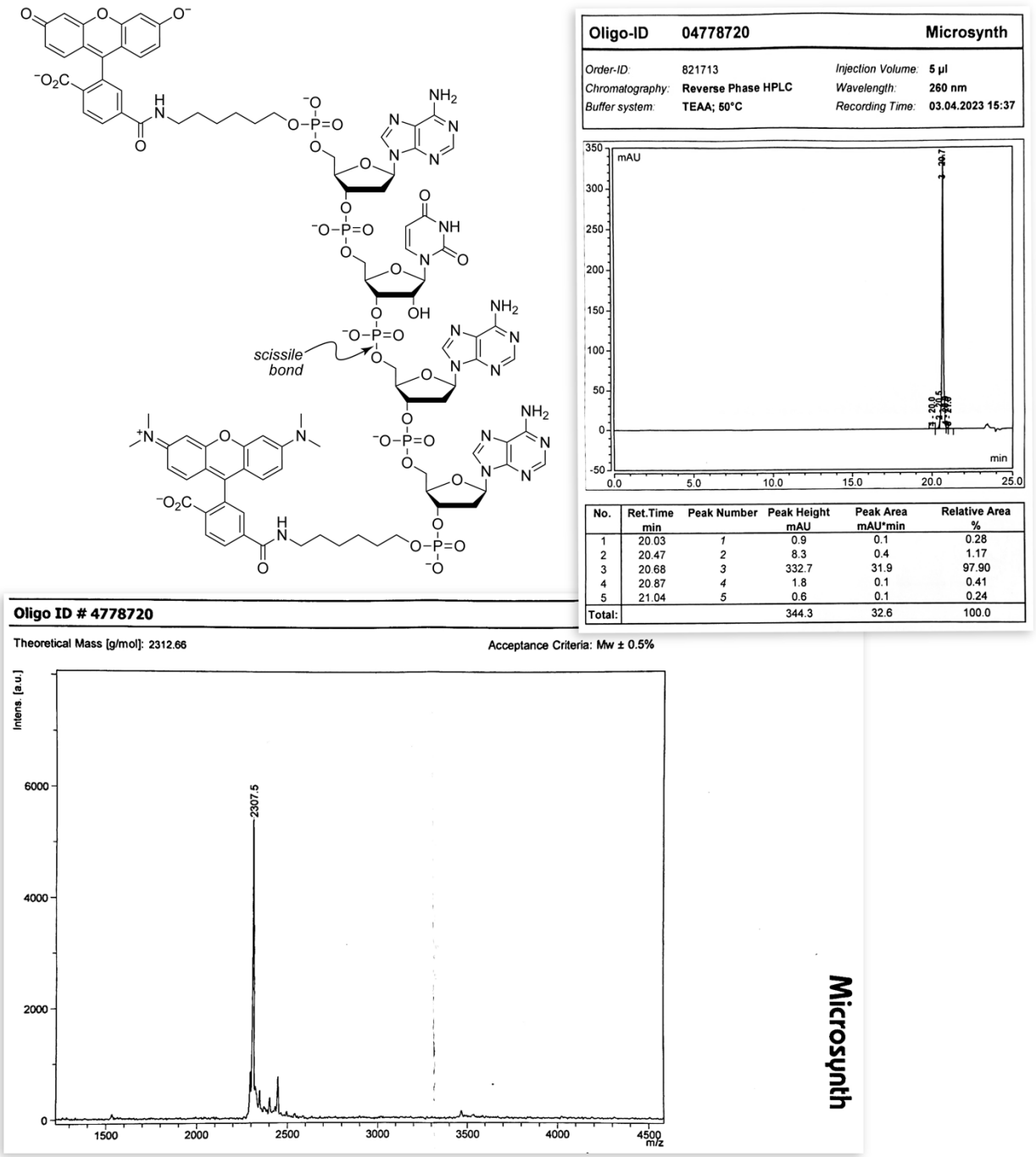


FIGURE S6 Structure of 6-FAM–dArU(dA)$_2$–6-TAMRA, which is the fluorogenic substrate for assays of ribonucleolytic activity and has only one scissile phosphodiester bond. The cleavage of that bond increases the fluorescence of the FAM moiety with λ_{ex} = 493 ± 5 nm and λ_{em} = 515 ± 5 nm. Insets: reversed-phase HPLC chromatogram (top) and MALDI–TOF mass spectrum (bottom) of synthetic 6-FAM–dArU(dA)$_2$–6-TAMRA.

REFERENCES

Anand K, Ziebuhr J, Wadhwani P, Mesters JR, Hilgenfeld R. Coronavirus Main Proteinase (3CL^{pro}) Structure: Basis for Design of Anti-SARS Drugs. Science 2003;300:1763–7.

Kneller DW, Phillips G, Weiss KL, Pant S, Zhang Q, O'Neill HM, et al. Unusual zwitterionic catalytic site of SARS-CoV-2 main protease revealed by neutron crystallography. J Biol Chem 2020;295:17365–73.

Koudelka T, Boger J, Henkel A, Schönherr R, Krantz S, Fuchs S, et al. N-Terminomics for the Identification of In Vitro Substrates and Cleavage Site Specificity of the SARS-CoV-2 Main Protease. Proteomics 2021;21:e2000246.

Ramos-Guzmán CA, Ruiz-Pernía JJ, Tuñón I. Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods. ACS Catal 2020;10:12544–54.

Primer: Ribonucleolytic Activity of the Uncleaved and Cleaved Zymogen

1 Defining the reaction system

We have a ribonuclease 1 (RNase 1) zymogen of concentration [Z] whose active site is occluded by a peptidic linker that is recog-nized and cleaved by $3CL^{pro}$. We let U represent the zymogen in its unactivated, pre-cleaved state and A represent the activated zymogen after linker cleavage. If Z represents the total population of zymogen regardless of its cleavage state, then by mass balance,

$$[\mathrm{U}] + [\mathrm{A}] = [\mathrm{Z}] \tag{S1}$$

Both the unactivated and activated zymogen are capable of cleaving a generic RNA substrate S to yield a product P, albeit with different catalytic efficiencies:

$$\mathrm{U} + \mathrm{S} \rightarrow \mathrm{U} + \mathrm{P} \tag{S2}$$

$$\mathrm{A} + \mathrm{S} \rightarrow \mathrm{A} + \mathrm{P} \tag{S3}$$

The unactivated zymogen hydrolyzes the substrate with catalytic efficiency $(k_{cat}/K_M)_U$, and the activated zymogen correspondingly hydrolyzes with efficiency $(k_{cat}/K_M)_A$.

Letting R represent $3CL^{pro}$, the reaction equation for zymogen cleavage is:

$$\mathrm{R} + \mathrm{U} \rightarrow \mathrm{R} + \mathrm{A} \tag{S4}$$

The protease likewise has a catalytic efficiency $(k_{cat}/K_M)_R$ for cleavage of the zymogen linker. In the context of the protease cleavage reaction, the unactivated and activated zymogen are substrate and product, respectively, rather than catalysts.

2 Modeling the time evolution of zymogen activation

Proteolytic activation of the zymogen follows standard Michaelis-Menten kinetics, and the initial rate v_R of zymogen activation represented by chemical reaction S4 is expressed by

$$v_R = \frac{k_{cat,R}[R][U]}{K_{M,R} + [U]} \tag{S5}$$

where $k_{cat,R}$ is the rate constant and $K_{M,R}$ is the Michaelis constant for proteolytic cleavage. We make our first simplifying assumption, the validity of which will be justified shortly:

Assumption 1: the concentration of unactivated zymogen is always much less than $K_{M,R}$

By invoking $[U] << K_{M,R}$, equation S5 can be approximated and solved for [U] as a function of time t:

$$v_R = -\frac{d[U]}{dt} = \left(\frac{k_{cat}}{K_M}\right)_R [R][U] = r[U]$$

where we define the pseudo-first order rate constant $r = \left(\frac{k_{cat}}{K_M}\right)_R [R]$. Then:

$$\frac{d[U]}{[U]} = -r\,dt$$

$$\int \frac{1}{[U]}\,d[U] = -r\int dt$$

$$\ln[U] + B = -rt + C$$

After consolidating the constants of integration B and C and cleaning up the equation, we are left with:

$$[U] = De^{-rt}$$

for some constant D. With equation S1, we can express the time-dependent concentration of activated zymogen by

$$[\mathrm{A}] = [\mathrm{Z}] - De^{-rt}$$

Applying the initial condition that $[\mathrm{A}] = 0$ at $t = 0$, we find that

$$[\mathrm{U}] = [\mathrm{Z}]e^{-rt} \tag{S6}$$

$$[\mathrm{A}] = [\mathrm{Z}]\left(1 - e^{-rt}\right) \tag{S7}$$

We note that the assumption $[\mathrm{U}] << K_{\mathrm{M,R}}$ is dependent on the protease of interest and the identity of the zymogen linker. In our 3CL$^{\mathrm{pro}}$-directed zymogen, we have installed the linker sequence ATLQSGNA. Our lab recently characterized 3CL$^{\mathrm{pro}}$ cleavage of the peptidic substrate R–E(EDANS)–ATLQSGNA–K(DABCYL)–R and found K_{M}=229 μM (at 25 °C in 50 mM HEPES, pH 7.5). It is likely that the zymogen linker (installed within a protein) is a poorer substrate than the analogous free peptide, thus $K_{\mathrm{M,R}}$ is believed to be higher for cleavage of the zymogen linker. The zymogen is utilized at concentrations well below this (e.g., high picomolar to low nanomolar), as it is itself being used as a catalyst in the cleavage of an RNA substrate. As a result, the approximation $[\mathrm{U}] << K_{\mathrm{M,R}}$ should be valid.

3 Characterizing ribonucleolytic activity over time

Both the unactivated and activated zymogen follow Michaelis-Menten kinetics in their cleavage of an RNA substrate (chemical reactions S2 and S3):

$$v_{\mathrm{U}} = \frac{k_{\mathrm{cat,U}}[\mathrm{U}][\mathrm{S}]}{K_{\mathrm{M,U}} + [\mathrm{S}]}$$

$$v_{\mathrm{A}} = \frac{k_{\mathrm{cat,A}}[\mathrm{A}][\mathrm{S}]}{K_{\mathrm{M,A}} + [\mathrm{S}]}$$

We now make our second simplifying assumption:

Assumption 2: the concentration of RNA substrate is always much less than both $K_{\mathrm{M,U}}$ and $K_{\mathrm{M,A}}$

This assumption can be rendered valid by the design of the zymogen activity assay, namely the initial substrate concentration $[\mathrm{S}]_0$ that we choose to run the assay at. The K_{M} for cleavage of the short oligonucleotide substrate dArU$(\mathrm{dA})_2$ by wild-type RNase A (the bovine homologue of the zymogen's progenitor human RNase 1, 68% sequence identity) is ~22 μM. Typical ribonucleolytic activity assays are conducted with substrate in the mid- to high-nanomolar concentration range; thus, this assumption is valid.

Applying Assumption 2 results in:

$$v_{\mathrm{U}} = \left(\frac{k_{\mathrm{cat}}}{K_{\mathrm{M}}}\right)_{\mathrm{U}}[\mathrm{U}][\mathrm{S}] = c_{\mathrm{U}}[\mathrm{U}][\mathrm{S}]$$

$$v_{\mathrm{A}} = \left(\frac{k_{\mathrm{cat}}}{K_{\mathrm{M}}}\right)_{\mathrm{A}}[\mathrm{A}][\mathrm{S}] = c_{\mathrm{A}}[\mathrm{A}][\mathrm{S}]$$

$$v_{\mathrm{T}} = -\frac{\mathrm{d}[\mathrm{S}]}{\mathrm{d}t} = v_{\mathrm{U}} + v_{\mathrm{A}} = \left(c_{\mathrm{U}}[\mathrm{U}] + c_{\mathrm{A}}[\mathrm{A}]\right)[\mathrm{S}]$$

For notational simplicity, we denote the catalytic efficiencies of the unactivated and activated zymogen by $c_{\mathrm{U}} = \left(\frac{k_{\mathrm{cat}}}{K_{\mathrm{M}}}\right)_{\mathrm{U}}$ and $c_{\mathrm{A}} = \left(\frac{k_{\mathrm{cat}}}{K_{\mathrm{M}}}\right)_{\mathrm{A}}$, respectively. As [U] and [A] are defined by equations S6 and S7, we can express the rate of RNA turnover as:

$$\frac{\mathrm{d}[\mathrm{S}]}{\mathrm{d}t} = -[\mathrm{Z}]\left(c_{\mathrm{U}}e^{-rt} + c_{\mathrm{A}}\left(1 - e^{-rt}\right)\right)[\mathrm{S}] \tag{S8}$$

Note that if $[\mathrm{R}] = 0$ (*i.e.*, no protease is present to cleave the zymogen), then equation S8 simplifies to the general expression for pseudo-first order enzyme kinetics, as would be expected when the zymogen remains unactivated at all times. Rearranging terms and integrating results in:

$$\int \frac{1}{[\mathrm{S}]}\,\mathrm{d}[\mathrm{S}] = -[\mathrm{Z}]c_{\mathrm{U}}\int e^{-rt}\,\mathrm{d}t - [\mathrm{Z}]c_{\mathrm{A}}\int\left(1 - e^{-rt}\right)\mathrm{d}t$$

$$\ln[\mathrm{S}] + J = \left(\frac{[\mathrm{Z}]}{r}\right)c_{\mathrm{U}}e^{-rt} - [\mathrm{Z}]c_{\mathrm{A}}t - \left(\frac{[\mathrm{Z}]}{r}\right)c_{\mathrm{A}}e^{-rt} + K$$

for some constants of integration J and K. Note that this expression for [S] requires that $r \neq 0$, which is to say that the expression is only applicable when the protease is present (for if [R] = 0, then $r = 0$). After consolidating the constants of integration and cleaning up the equation, we are left with

$$[\mathrm{S}] = L \exp\left(\frac{[\mathrm{Z}]}{r}\left(c_\mathrm{U} - c_\mathrm{A}\right)e^{-rt} - [\mathrm{Z}]c_\mathrm{A}t\right) \tag{S9}$$

for some constant L. If we apply the initial condition that at $[\mathrm{S}] = [\mathrm{S}]_0$ at $t = 0$, we learn that

$$L = [\mathrm{S}]_0 \exp\left(-\frac{[\mathrm{Z}]}{r}\left(c_\mathrm{U} - c_\mathrm{A}\right)\right)$$

Inserting this into equation S9 results in:

$$\begin{aligned}
[\mathrm{S}] &= [\mathrm{S}]_0 \exp\left(-\frac{[\mathrm{Z}]}{r}\left(c_\mathrm{U} - c_\mathrm{A}\right)\right)\exp\left(\frac{[\mathrm{Z}]}{r}\left(c_\mathrm{U} - c_\mathrm{A}\right)e^{-rt} - [\mathrm{Z}]c_\mathrm{A}t\right) \\
&= [\mathrm{S}]_0 \exp\left(\frac{[\mathrm{Z}]}{r}\left(c_\mathrm{U} - c_\mathrm{A}\right)\left(e^{-rt} - 1\right) - [\mathrm{Z}]c_\mathrm{A}t\right) \\
&= [\mathrm{S}]_0 \exp\left(\frac{[\mathrm{Z}]}{r}\left(c_\mathrm{A} - c_\mathrm{U}\right)\left(1 - e^{-rt}\right) - [\mathrm{Z}]c_\mathrm{A}t\right)
\end{aligned}$$

Finally, if we make a common assumption of enzyme assays that the substrate concentration greatly exceeds that of the enzyme (i.e., [S] >> [Z]), we can neglect the concentration of zymogen-bound substrate and say that the product concentration is $[\mathrm{P}] = [\mathrm{S}]_0 - [\mathrm{S}]$. Thus,

$$[\mathrm{P}] = [\mathrm{S}]_0\left(1 - \exp\left(\frac{[\mathrm{Z}]}{r}\left(c_\mathrm{A} - c_\mathrm{U}\right)\left(1 - e^{-rt}\right) - [\mathrm{Z}]c_\mathrm{A}t\right)\right) \tag{S10}$$

4 Monitoring ribonuclease activity by fluorescence

We monitor ribonucleolytic activity using a FRET-tagged, chimeric oligonucleotide substrate 6-FAM–dArU$(\mathrm{dA})_2$–6-TAMRA in which the core tetranucleotide has one scissile bond (rU↓dA); substrate cleavage liberates the donor FAM from the quencher TAMRA, inducing fluorescence and enabling monitoring product formation. The relationship between fluorescence intensity F and product concentration [P] is linear:

$$\begin{aligned}
F &= \alpha[\mathrm{P}] + F_0 \\
\Rightarrow [\mathrm{P}] &= \alpha^{-1}\left(F - F_0\right)
\end{aligned}$$

where α is a proportionality constant (dictated by experimental parameters such as instrument gain) and F_0 is the background fluorescence intensity. Given that the maximum concentration of product $[\mathrm{P}]_\mathrm{max} = [\mathrm{S}]_0$, there is a corresponding maximum fluorescence intensity:

$$\begin{aligned}
F_\mathrm{max} &= \alpha[\mathrm{P}]_\mathrm{max} + F_0 = \alpha[\mathrm{S}]_0 + F_0 \\
\Rightarrow [\mathrm{S}]_0 &= \alpha^{-1}\left(F_\mathrm{max} - F_0\right)
\end{aligned}$$

Inserting the above expressions into equation S10 allows us to model product fluorescence as a function of time:

$$F = F_0 + \left(F_\mathrm{max} - F_0\right)\left(1 - \exp\left(\frac{[\mathrm{Z}]}{r}\left(c_\mathrm{A} - c_\mathrm{U}\right)\left(1 - e^{-rt}\right) - [\mathrm{Z}]c_\mathrm{A}t\right)\right) \tag{S11}$$

5 Summary

We have an RNase 1 zymogen of concentration [Z] whose active site is occluded by a peptidic linker that is recognized and cleaved by $3\mathrm{CL}^\mathrm{pro}$. If at $t = 0$ we mix the zymogen with $3\mathrm{CL}^\mathrm{pro}$, then the concentration of uncleaved/unactivated and cleaved/activated zymogen at any time t is determined by

$$[\mathrm{U}] = [\mathrm{Z}]e^{-rt} \tag{S12}$$

$$[\mathrm{A}] = [\mathrm{Z}]\left(1 - e^{-rt}\right) \tag{S13}$$

where $r = \left(\frac{k_\mathrm{cat}}{K_\mathrm{M}}\right)_\mathrm{R}[\mathrm{R}]$.

If, at the same time that we mix the zymogen and $3\mathrm{CL}^\mathrm{pro}$, we mix in the oligonucleotide RNase 1 substrate 6-FAM–dArU$(\mathrm{dA})_2$–6-TAMRA, then the fluorescence intensity of the product is governed by

$$F = F_0 + \left(F_\mathrm{max} - F_0\right)\left(1 - \exp\left(\frac{[\mathrm{Z}]}{r}\left(c_\mathrm{A} - c_\mathrm{U}\right)\left(1 - e^{-rt}\right) - [\mathrm{Z}]c_\mathrm{A}t\right)\right) \tag{S14}$$